Supplementary Information

Chromatin Immunoprecipitation - Indirect peaks Highlight Functional Long-Range Interactions among Insulator Proteins and RNAPII Pausing

Jun Liang[+*], Laurent Lacroix[+*], Adrien Gamot[+], Suresh Cuddapah^, Sophie Queille[+], Priscillia Lhoumaud[+], Pierre Lepetit[+], Pascal G.P. Martin[+], Jutta Vogelmann[!], Franck Court[+], Magali Hennion[+], Gaël Micas[+], Serge Urbach[$], Olivier Bouchez`, Marcelo Nöllmann[!], Keji Zhao^, Eldon Emberly[#] and Olivier Cuvier[+@]

Experimental procedures

DNA constructs, Yeast two hybrid

Full-length *beaf32* or *cp190* were obtained by PCR from cDNA subsequently cloned into the pDONR221. For mutagenesis of *beaf32*, PCR-mediated mutagenesis was performed using pDONR221-*beaf* and primers that contain mismatches in location of 422-425, and 435-436 of the full-length *beaf32* gene. Synthetic genes were obtained by gene synthesis after designing silent mutations replacing all available degenerated codons corresponding to the first 800 nucleotides of *beaf32*. To express mut- or WT- Beaf at low levels, synthetic genes were cloned into 'pWG' obtained by removal of the NotI EcoRI region of the relatively weak ubiquitin promoter of pUWG (Akbari et al., 2009). For selection, the *neomycin* gene of 'pEGFP' was further inserted through ligation-mediated PCR into the NdeI site of pWG. Reporter constructs-used for ChIP or for expression analyses were generated by inserting a synthetic FspI/NotI fragment corresponding to the *sop* gene (-500 from TSS (including the direct Beaf32 peak) until the end of the second exon) followed by *hsp27* (-500 from TSS (harboring the direct GAF site) to the end of the second exon of *hsp27* followed by in-frame fusion with g*fp*) into pMK-RQ digested with SmaI/NotI. The GAF sites were mutagenized through PCR-mediated mutagenesis.

Reporter constructs were transiently transfected into cells expressing WT- or mut- Beaf (stable cell line) and RNA or DNA were extracted after 3 days for analysis by RTqPCR or ChIP. All vectors were checked through DNA sequencing using oligos available upon request. For co-immunoprecipitation experiments, the synthetic *beaf* (WT or mutant) genes were fused to the GFP cassette using available vectors (Gateway). For yeast 2 hybrid, bait constructs and target constructs were cloned into pDest22 and pDest32 to express fusion proteins containing the yeast. Growth of yeast strain MaV203 expressing WT- or mutant -Beaf and CP190 proteins in various combinations. Yeast cells were plated on nonselective (+Ura) and selective (-Ura) incubated at 30°C for 2 days to examine the phenotypes on SD media. *ura3* reporter genes was used in the yeast two-hybrid assays (Invitrogen).

Antibodies

Anti-Beaf32/CP190/dCTCF antibodies were generated in rabbit using KLH-coupled with the last 15AA of the C-terminal domain and were then affinity-purified as previously described (Cuvier and Hirano, 2003). Antibody specificity was confirmed by western blotting of nuclear extracts and by competition with peptides (Fig. S2).

Nuclear Extracts , Immunoprecipitation , peptide mapping

Nuclear extracts were prepared by extracting nuclei on ice in 1 mL of 10 mM HEPES (pH 7.6), 350 mM KCl, 0.1 mM EDTA, 3 mM $MgCl_2$, 1 mM DTT, 10% glycerol, 0.2% Trasylol, 0.2 mM PMSF, followed by centrifugation at 54000rpm for 60 min at 4°C. The supernatant was stored in buffer (25 mM HEPES [pH 7.6], 250 mM KCl, 0.1 mM EDTA, 1 mM DTT, 10% glycerol, 0.1 mM PMSF). All steps were performed at 4°C. Nuclear extracts were clarified by centrifugation

at 16,000 × g for 10 min. Protein A Sepharose (200 μL) was washed three times in PBS with 0.05% NP40, Rabbit polyclonal anti-Beaf32 serum (50 μL) or preimmune serum was added and incubated overnight. Beads were washed three times with 1x PBS + 0.05% NP40, then mixed with 750ul nuclear extracts overnight at 4°C. The beads were loaded into chromatography column (Bio-rad), washed with 2 mL PBS-0.05% NP40 and then the same buffer + 0.4 M KCl. The beads were aliquoted into 22 tubes and incubated with 18 individual peptide at the concentration of 0.4 mg/mL and 2 control peptides for 3 hours at 4 °C. Supernatants were collected after spinning down the beads for 5 minutes at 10000 rpm. Samples were boiled in protein loading buffer for 5 min and run the protein gel. After the protein was transferred to nitrocellulose by diffusion transfer as previously described (Cuvier et al., 2002), blots were probed with anti-Beaf32 at 1:2500, anti-CP190 at 1:5000, anti-H3 at 1:10000, followed by horseradish-peroxidase-conjugated donkey anti-rabbit (1:10,000) (GE Healthcare). Signals were developed using an ECL Plus detection kit (GE Healthcare) and FUJIFILM Luminescent Image Analyzer LAS-4000 for images acquisition.

S2 cells, RNAi, stable transfections, RNA sequencing, Gene expression analysis

Drosophila S2 cells were maintained between $1.2x10^6$ cells/ml and $2.4x10^6$cells/ml in Schneider *Drosophila* medium (Invitrogen) supplemented with 10% Fetal Bovine Serum (FBS) and 1% penicillin/streptomycin (GIBCO). Stable transfections of $3.6x10^6$ S2 cells were done according to handbook using 10 μg pUWG-neo or pWG-neo serial vectors carrying *neomycin* resistance gene precipitated with calcium phosphate. G418 resistant colonies appeared after about 10 days under selection. The cells were grown as a mixed culture. For RNAi-mediated depletion, T7-driven synthesis (Fermentas TranscriptAid™ T7 High Yield Transcription Kit) of double-

stranded RNAs (dsRNA) specific for *beaf32*, *dctcf* or control/mock dsRNA (of similar sizes) checked for potential off-target effects using NCBI primer designing tool and dsCheck (http://dsCheck.RNAi.jp/) was used. 400 μg of dsRNA were added to 36 millions cells in 10 mL media without FBS. Cells were incubated for 2h at 25°C and 20mL of media with FBS were added. After incubation for 5 days, the cells were harvested, followed by RNA extraction (Qiagen Rneasy; deep-seq and RT-qPCR) or formaldehyde crosslinking for ChIP analysis. Gene expression was measured by real-time PCR analysis using cDNAs prepared from S2 control, Beaf32- or dCTCF- -depleted cells, for the indicated genes (primers are available upon requests). Quality of mRNA levels were controlled with Experion (Biorad) and quantified in parallel with at least five different concentrations of cDNA and genomic DNA for standard curves using Biorad iQ SYBR Green Supermix in an Eppendorf Realplex. RTqPCR data were analyzed using the relative quantification component of pyQPCR software developed by M. Hennion and T. Gastine accessible at (http://pyqpcr.sourceforge.net/) establishing high (95%; confidence interval) PCR efficiency. Depletions of Beaf32 or dCTCF were verifired by RTqPCR (~95%). Beaf32 depletion impairs the expression of both Beaf32A/B isoforms as measured by RTqPCR using primers spanning exon-exon junctions (Fig. S4G). To measure 32A expression after expression of mut- or WT-Beaf compared to mock-depleted cells, RNASeq reads spanning the junction of exon 1 and 2 of 32A specifically were counted using DiffSplice (Hu et al., 2013). Samples were analyzed as replicates by RNA sequencing (HiSeq2000; Illumina) at the local INRA plateform (https://genomique.genotoul.fr/) or through BGI. All RNASeq analyses (RNA-seq; Hi-Seq2000; Illumina) were performed using TopHat or the Burrows Wheeler Alignment tool (BWA) software (default parameters) on genome annotations of release 5.41 of Flybase for parsing, HTseq for counting the reads and DEGseq package to identify differentially expressed

genes (p-value<0.001) from 3 independent replicates. Note that of these DE genes, >80% showed changes in expression > 1.6 fold as compared to control. Also, all intersection analyses between DE genes and additional genomic features (including with direct or indirect peaks) were further verified by applying such minimal fold change in expression (Figure S4I; column "3b", Figure 5B). Proportional Venn diagrams were plotted using http://bioinforx.com. Statistical enrichments according to the distance from ChIP-Seq peaks (Beaf32, dCTCF or GAF) were performed by intersecting groups of genes within sliding (200 bp) windows of 2,500 bp with the indicated feature (e.g. long-range targets with or without ChIP-indirect peak in their promoter) compared to total genes within each window using fisher's exact tests using 2X2 contingency tables. p.values were corrected using Benjamini & Hochberg multiple testing corrections. RNASeq in RNAi depleted cells stably transfected with mutant- or WT- Beaf were performed three independent times. Heat maps (Figure 5) were obtained by intersecting the lists of long-range targets within < 40 kbp from a direct Beaf32 peak with the indicated genomic features starting from gene lists of differentially expressed genes harboring ChIP-indirect peaks of 630/693/623/561 in sizes (obtained from the 3 independent RNASeq).

3C assays

Chimera (religation products) in 3C samples were measured by qPCR using Taqman technology in Beaf32-depleted, CP190-depleted or control S2 cells (Figure 3) and from Beaf32-depleted stable cell lines expressing mutant-Beaf (Figure S3). Following RNAi treatment, cells were cross-linked with 1%PFA for 10 min. and the reaction was stopped by addition of 125 mM glycine with 3x washes (PBS). 2 10^7 cells were used to prepare each 3C DNA sample. Briefly, 3C DNA was prepared as follow: Cells (2 10^7) were resuspended and washed in Hind III buffer

then lyzed after addition of 0.2% SDS. After SDS neutralization with Triton X100 (1.2% final), digestion was performed at 37°C overnight with 2000u of HindIII (NEB R0104M). Digestion was stopped with SDS (1% final) and 65°C incubation (20 min). Digestion efficiency was evaluated using primers covering 12 HindIII sites within the *Tsp39D/crc/mio* locus. Similar mean efficiencies of digestion (55-90%) were obtained for each sites between all samples. After dilution in ligation buffer (~1/20 dilution), SDS was trapped with Triton X100 (1% final) and ligation was performed at 16°C for 3h30 in presence of 210u of T4 DNA ligase (Fermentas) and ATP (1mM). Reaction was stopped with EDTA and DNA recovered by Phenol/chloroform and chloroform extractions followed by ethanol precipitation. DNA was further digested with EcoRI (that does not cut within the expected chimera products). DNA concentration was evaluated with qPCR experiment with the SYBR-Green technology using 3 couples of control primers from each locus, covering regions where no EcoRI or HindIII site was present. Amount of 3C DNA recover were typically in the 50-60μg range for 2 10^7 cells.

Couples of primers used for chimera quantification were validated on random ligation product from BAC spanning the *tsp39D/crc/mio* locus (BACR34H23 and BACR05D08) with SYBR-green and Taqman-MGB qPCR approaches (Court et al., 2011). Relative frequencies of chimera were calculated relative to the total DNA amount assessed using the 3 control regions (within the same locus) used for concentration determination. Similar data were obtained for quantification relative to the anchor site in control and depleted cells. PCR efficiencies of all the primer couples are between 97 and 108% with the Taqman-MGB probe. All primers were also independently tested with SYBR-Green based qPCR. Chimera frequencies were measured in triplicate using 20-150ng of 3C treated DNA for each data point. 3C Data presented in Figure 3 and S3 represent the average of 2-3 independent 3C DNA preparation for which chimera frequencies have been

measured independently 3 times in triplicate each time. Variations of the signal over the indirect peaks or nearby regions were statistically tested by Student test between cells Beaf32- or CP190-depleted compared to mock-depleted cells (p-values of 0.03 and 0.01, respectively). Error bars represent the standard error to the mean arising from the variability between the independent 3C DNA preparations. Frequencies were calculated using the $2^{-\Delta C_t}$ formula relative to the total amount of DNA. qPCR were performed on Viia7 (Life Technology) and Realplex (Eppendorf) apparatus and Cq evaluated using the relative threshold (Life Technology) or CalQplex (Eppendorf) algorithm. For detection of chimera, all primers were 'reverse' primers with respect to standard chromosome coordinates. Primers and probes sequences are available upon request.

Statistical analyses of Hi-C data by aggregation

Genome-wide statistical analyses were performed through aggregation plots (Jee et al., 2011) using Hi-C data generated with the frequent cutter Dpn II (Sexton et al., 2012). For statistical analyses of interactions, we started from the previously characterized data set of parsed reads called 's0' that were shown to provide with high-quality contact maps at 10-20 kbp resolution (Sexton et al., 2012). The genomic features - presence or absence of Beaf32, dCTCF, GAF or CP190 peaks and/or high/low RNAPII pausing site (+ 40 after TSSs) - were mapped with respect to the first restriction fend as previously done (Sexton et al., 2012) to generate a matrix of genomic fends. To measure interactions between "far *cis*" fends (distant fends on the same chromosome arms), a minimal distance of 15 kbp was taken between such fends due to possible limitations of quantitative coverage below a few kbp (Sexton et al., 2012)(not shown). Far cis interacting fends were counted providing a binary contact matrix for each fragment end pair as previously (Sexton et al., 2012). For inspection of contact maps, the binary interaction matrix

was previously normalized over bins of 10-20 kbp (Sexton et al., 2012) whose sizes depend on sequence coverage. Further aggregation of the data largely increases read coverage (due to the genome-wide alignment of the signal over the indicated sites), which was based on the following modifications. Interactions were measured between all fends harboring a given genomic feature (e.g. all fends harboring a direct peak) and the distant/paired fends harboring the second feature of interest (e.g. one of the thousands of indirect peaks localized > 15 kbp away). The aggregation of the signal was performed over this later paired fend that defined the position 0 of the x-axis. Interactions were then normalized over the number of genomic fends for every 500 bp window/bin. The resulting amplification of the signal (e.g. over approximately 2,700 indirect peaks) therefore provided with an averaged interaction profile every 500 bp from the site of interest (+/- 5 kbp) independently of the distribution of genomic restriction fends. We independently verified that the distribution of Dpn II fends was similar for bins harboring the specific genomic features of interest as compared to control bins. Interaction profiles were then oriented such that the distant sites (e.g. the direct peaks) were always on the left of the plot. For internal controls of interactions, the same procedure was applied by aggregating the profiles over neighbor bins localized at a constant distance (+/- 3; 3.5; up to 9 kbp) thus generating 26 control interaction profiles as measured for each pair of direct-indirect peaks (2,700 times). Such control plots were also aggregated over each control bin providing with 26 control aggregation plots. The signal from the 26 neighbor windows generated a background signal (shadowed region of the plots) reflecting the interaction levels within the same chromatin contexts. The enrichment in long-range interactions was then calculated as the ratio of the interactions over the original position 0 (the indirect peaks) over the background interaction levels from internal controls (neighbor bins) as calculated for every successive 500 bp window. Note that similar results were

obtained if background signal was defined by measuring interactions levels from internal controls selected by counting a fixed number of restriction fends from the bin of interest. Frequencies of interactions were further tested through the following procedures. First, interactions were measured for 3 complementary subgroups of genomic direct/indirect peaks as an estimate of the variation of the interaction levels. We also selected 3 complementary subsets of paired fends (from the interacting matrix) that were used to generate error bars of the measures. Second, bootstrapping was utilized to verify the reproducibility of the enrichment in interactions (not shown). Third, Shapiro-Wilk tests and QQ-plots were used to validate the normal distribution of the normalized interaction levels (distribution of interaction levels over 2,700 x 26 control bins) providing with 95% confidential intervals. P-values were then calculated based on such normal distribution using the 'pnorm' function of R as the probability of obtaining the levels of interactions at the levels found for the genomic feature of interest. An additional constrains for measures of long-range interactions were also applied including an upper limit of 60 kbp (up to 200 kbp; Supplem. Fig. S2) for the maximal distance separating paired fends, corresponding to the averaged sizes of topologically activating domains (approximately 60 kbp; Sexton et al., 2012; Hou et al., 2012). For long-range interactions profiles in the context of TAD borders, direct / indirect pairs were selected from two successive domains separated by a topological border (see Sexton et al., 2012 for a list; "Active-Active"), as compared to pairs found within the same TAD. For comparison, the two profiles were generated using the same lower/upper limits for distances (i.e. 15 kbp < d < 60 kbp) between paired fends.

Fluorescence anisotropy

Anisotropy measurements were carried out using a Tecan Safire II micro plate reader Fluorimeter and a Corning 384 well flat bottom microplate. Measurements were carried out in

30mM Tris buffer pH 7.5, 0,01 mg/mL BSA, 0,04% Tween20, 100mM NaCl, 20 μM ZnSO4, 5 mM Mercaptoethanol in a 384 Well Low Flange Black Flat Bottom in a total volume of 60μL. DNA binding studies were performed adding increasing amounts (0-200 nM) of purified proteins to a 2.5 nM concentrated 5` labelled Cy3B or atto-655 58bp oligonucleotides. The sequence of the dsDNA fragements used was: specific DNA sequence: AGA AGT CAG CGC GAT AGC ATC GAT ATT TTC GTG ACA CGC TTG TCA TCC GAT AGG TAG T. ; non-specific DNA sequence: GGA CAG GTA TTG TGC CAT ACT GAC CAC ATC GTC TTG GTC TAT AAG CTC CAC GAC ATC C. Equilibrium binding data did not fit well to a single binding site model (data not shown). Thus, they were fitted with a Hill model with the apparent dissociation constant (K_d) and cooperativity (n) as free parameters.

ChIP-Seq, ChIP-Ind, Intersection analyses

For Beaf32 and CP190 ChIP-Seq analyses, the ChIP samples were processed to construct the sequencing libraries and sequenced as previously (Barski et al., 2007). To generate genome-wide profiles from sequencing reads, first reads were shifted by an average fragment size that was determined from the spatial correlation function between (+) and (-) reads. Genome-wide profiles were then generated from the read density using a Gaussian smoothing procedure. We computed the average fragment size and variance by calculating the spatial correlation function between the + and – reads (25 bp reads from either the 5' (+) or 3' (-) end of each fragment). This was done by making a histogram of the distances between all + and – reads across the genome (Fig. S1). The first peak in the histogram represents the average fragment size and its width the uncertainty in position. To make a profile, each read's contribution to the profile is given by a Gaussian density centered on the read with a fixed standard deviation. The total score

at a given location is the sum of all the read densities in the local neighborhood. For the genome-wide profile, we used an average shift of 45 bp and a standard deviation of 50 bp. The distribution of peaks (Figure 1A, S1) was analyzed with respect to each of *Drosophila* TSSs to score ChIP-Seq reads (+/- 250bp from TSS; 10bp bins) using raw Chip-seq data. Peaks/TSS were then ranked according to the density of reads and to the average profile of Beaf32 binding for given read density (Fig. S1). Plots representing the % any genomic feature (e.g. 'CGATA' consensus or CP190/NELF binding etc.) were generated according to the corresponding peak density. Given the under-enrichment of CGATAs within 'small peaks' (>10 and <25 read counts) that otherwise define a specific ChIP profile along TSS (Fig. S1), we sorted out specific ChIP-indirect peaks corresponding to promoters with no CGATA (>900 peaks in promoters), and direct peaks >25 with >3 CGATAs (4,120 in promoters). The position of the maximal values of every peak was then calculated (read counts in 10bp bins). Adjusted p.values utilized Benjamini & Hochberg correction. Note that ChIP-indirect peaks of Beaf32 were not enriched in DNA binding consensus related to Beaf32, to the 32A isoform or to DREF (Figure S1). ChIP-indirect peaks of Beaf32 in *Drosophila* embryos (Negre et al., 2010) were analyzed with rMAT package. Binding profiles were obtained following the same averaging procedure used to analyze ChIP-Seq data from S2 cells. Enrichment in CGATA-less Beaf32 peaks with respect to the distribution of CP190 sites (Figure 1D) was obtained by counting the number of ChIP-indirect peaks for each interval and as a function of the distances to direct Beaf32 peaks followed by normalization to the total number of CP190 sites. Variations in CP190 binding were detected between stably transfected cells expressing mut- or WT- Beaf using rMAT for peak detection normalized to input control. Variations in CP190 binding levels were obtained by measuring the differences in ChIP-Seq reads between cells expressing mut- compared to WT- Beaf, normalized

to the total read counts in the corresponding 100 bp window (CP190 peak identified by rMAT). For Figure S6B, 'new' CP190 peaks within the 0-100 bp window of genes and unique to cells expressing mutants were considered. To test the statistical enrichment of CP190 variations, the standard deviation in Log p-value was calculated using the numbers of CP190 peaks found in each interval (2 kbp) from two independent replicates. Impact of mutants was also verified in comparison with ChIP-Seq of CP190 from wild-type cells. Heat maps (Figure 6) were obtained by intersecting the number of peaks showing variations in CP190 levels (from two independent replicates) within the 0-100 bp window after the TSSs of genes, with the indicated genomic features: the corresponding numbers of genes are 791/805 (ChIP-Ind), 105/102 and 164/162 (dCTCF sites +/- ChIP-Ind), 186/193 and 382/371 (GAF sites +/- ChIP-Ind), 253/252 for ChIP-indirect peaks without dCTCF/GAF (47 of which were Su(Hw) binding sites; p-value~ 1e-7), 225/208 (long-range targets), 233/19 and 624/507 (highly paused genes +/- ChIP-Ind), 73/60 and 239/203 (active genes +/- ChIP-Ind). For detection of dCTCF, or GAF (direct)/(ChIP-Ind) peaks, data were obtained by analyzing ChIP from modencode / group of K. White (Negre et al., 2010), D. Gilmour (Gilchrist et al., 2008) and Renkawitz (Bartkuhn et al., 2009) respectively, following a similar procedure as for Beaf32. dCTCF ChIP-indirect peaks were defined by both read counting in promoter regions (+/-250bp) (dCTCF: direct > 10; ChIP-indirect peaks >4, <8;) after analyzing the peak threshold for statistical enrichment or not in previously defined dCTCF consensus (Bushey et al., 2009; Negre et al., 2010; Rhee and Pugh, 2011). Direct and/or ChIP-indirect peaks were then selected by taking into consideration the presence/absence of consensus within each subgroup pre-defined by read counts. Note that a low proportion of indirect peaks may harbor consensus sequences (see example in Table S4), which mostly depended on the threshold used to define indirect peaks. Promoters with < 2 reads were used as controls. For

GAF, direct/ChIP-indirect peaks were similarly defined based on ChIP-chip by enrichment of consensus from the direct peak group (Log2 enrichment > 5) in contrast to ChIP-indirect peaks (Log2 enrichment >2, <3) and controls (Log2 ~ 0). For intersection analyses, genes associated with overlapping sites between two distinct IBPs (e.g. between direct peaks of GAF (and/or dCTCF) and Beaf32) or for the same IBP (ChIP-indirect and direct peaks) were excluded including for the analyses of their regulation as 'long-range targets. RNA Pol II pausing profiles were obtained from our ChIP-Seq data in WT S2 cells (Gilchrist et al., 2010). The list of genes and their promoter features that were analyzed by ChIP of Beaf32 and CP190 is provided below:

gene	Beaf32 direct peak	Beaf32 - ChIP-indirect peak	GAF peak	dCTCF peak	CP190 peak	down-reg (mut/ WT)	up-reg (mut/WT)
CG10137	1	0	0	0	1	0	0
CG8026	1	0	0	0	1	0	0
mRpL54	1	0	0	0	0	0	0
CG3760	1	0	0	0	1	1	0
Cdc27	1	0	0	0	1	0	0
CG6673	1	0	0	0	0	0	0
Rpn1	1	0	0	0	1	0	0
Pgd	1	0	0	0	1	0	0
CG7995	1	0	0	1	1	0	0
CG33116	0	1	0	0	1	1	0
drpr	0	1	0	0	1	1	0
RpL18	0	1	0	0	1	0	1
RhoGDI	0	1	0	0	1	1	0
CG8029	0	1	0	1	1	1	0
fok	0	1	1	0	1	0	1
CG10527	0	1	1	0	1	1	0

CG15861	0	1	1	0	1	0	1
CG6776	0	1	1	0	1	1	0
ph-p	0	1	1	0	1	0	1
crc	0	1	1	1	1	0	1

Genome Browser

All genomic sequencing data including ChIP -Seq data of Beaf32, RNA- Seq upon Beaf-/dCTCF–depletion, RNA-Seq and ChIP-Seq of CP190 both in stably transfected cells expressing mutant/WT–Beaf, are accessible through NCBI or through the server http://insulators_chromosome-dynamics.biotoul.fr/IBPs

Contributions of Authors

-J.L. and L.L. contributed equally to the work. J.L. performed the experiments in Fig. 3, 4, 5, S3A-D, S4B-D, S6C, S7E under the supervision of O.C. -L.L. performed all 3C assays (Fig. 2; S2) and RTqPCR expression analyses (S4) with initial advising and help of additional O.C's lab members (F.C.; E. Guillou). A.G. performed all statistical analyses with O.C with help from E.E. S.C. and M.H. performed the ChIP of Beaf32 (Fig. 1) under the supervision of K.Z. and O.C. P.Lh. and P.M. did the ChIP of CP190 (Fig. 6, 7; S6) with J.L. J.L. and S.Q. performed dCTCF and CP190 depletion (Fig.S5). J.V. performed the the expression and *in vitro* binding assays using recombinant mutant/WT- Beaf under the supervision of M.N. and produced dCTCF antibodies to confirm depletions (S4E-F; not shown). O.B. supervised library constructions and Hi-Seq. S.U. performed the mass-spectrometric analysis (CP190 identification; Fig. 3B). A.G., G.M. and P.Le computed Hi-C data with O.C. (2C-D;S5I, S7D;7C) who designed and supervised all experiments and wrote the manuscript with corrections by all authors.

REFERENCES

Akbari, O.S., Oliver, D., Eyer, K., and Pai, C.Y. (2009). An Entry/Gateway cloning system for general expression of genes with molecular tags in *D. melanogaster*. BMC Cell Biol *10*, 8.

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Bartkuhn, M., Straub, T., Herold, M., Herrmann, M., Rathke, C., Saumweber, H., Gilfillan, G.D., Becker, P.B., and Renkawitz, R. (2009). Active promoters and insulators are marked by the centrosomal protein 190. Embo J.

Cattoni, D.I., Chara, O., Godefroy, C., Margeat, E., Trigueros, S., Milhiet, P.E., and Nollmann, M. (2013). SpoIIIE mechanism of directional translocation involves target search coupled to sequence-dependent motor stimulation. EMBO Rep *14*, 473-479.

Court, F., Baniol, M., Hagege, H., Petit, J.S., Lelay-Taha, M.N., Carbonell, F., Weber, M., Cathala, G., and Forne, T. (2011). Long-range chromatin interactions at the mouse Igf2/H19 locus reveal a novel paternally expressed long non-coding RNA. Nucleic Acids Res *39*, 5893-5906.

Cuvier, O., Hart, C.M., Kas, E., and Laemmli, U.K. (2002). Identification of a multicopy chromatin boundary element at the borders of silenced chromosomal domains. Chromosoma *110*, 519-531.

Cuvier, O., and Hirano, T. (2003). A role of topoisomerase II in linking DNA replication to chromosome condensation. J Cell Biol *160*, 645-655.

Gilchrist, D.A., Nechaev, S., Lee, C., Ghosh, S.K., Collins, J.B., Li, L., Gilmour, D.S., and Adelman, K. (2008). NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. Genes Dev *22*, 1921-1933.

Hu, Y., Huang, Y., Du, Y., Orellana, C.F., Singh, D., Johnson, A.R., Monroy, A., Kuan, P.F., Hammond, S.M., Makowski, L., *et al*. (2013). DiffSplice: the genome-wide detection of differential splicing events with RNA-seq. Nucleic Acids Res *41*, e39.

Jee, J., Rozowsky, J., Yip, K.Y., Lochovsky, L., Bjornson, R., Zhong, G., Zhang, Z., Fu, Y., Wang, J., Weng, Z., *et al*. (2011). ACT: aggregation and correlation toolbox for analyses of genome tracks. Bioinformatics *27*, 1152-1154.

Juang, B.H., and Rabiner, L.R. (1985). A Probabilistic Distance Measure for Hidden Markov-Models. At&T Technical Journal *64*, 391-408.

Negre, N., Brown, C.D., Shah, P.K., Kheradpour, P., Morrison, C.A., Henikoff, J.G., Feng, X., Ahmad, K., Russell, S., White, R.A., *et al*. (2010). A comprehensive map of insulator elements for the *Drosophila* genome. PLoS Genet *6*, e1000814.

Rhee, H.S., and Pugh, B.F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell *147*, 1408-1419.

SUPPLEMENTARY FIG. LEGENDS

Supplementary Figure S1 (related to Figure 1). High-resolution ChIP-Seq analysis highlights Beaf32 ChIP-indirect peaks.

A. Mapping the top (positive, '+' reads) and bottom (negative, '-' reads) strands separately on the genome provides high-resolution distribution profile of sequence reads for ChIP-Seq with anti-Beaf32 antibodies (left graph) as shown by plotting the distances between all 5'/+ and 3'/- reads (average distance ~82.5 bp). Note that a similar distribution of reads obtained from IgG controls (right graph) shows a largely reduced peak near this position due to the lack of immunoprecipitation from specific loci. The smaller peak near position 0 is also present in the IgG control due to reads in opposite orientation (as obtained from the same sonicated input material).

B. List of the two major consensus motifs found in the 3,411 Beaf32 peaks with their associated E-values as obtained by MEME (see Experimental procedures).

C. Graphs representing the distributions of Beaf32 direct peaks (top panel; >25 reads) and indirect peaks (middle panel, 10-25 reads / promoter) or of control sites defining background signal (<10 reads / promoter; bottom panel). The maximum value of each peak was taken to plot the averaged positioning of all peaks within a category (control, direct-, indirect- peaks) with respect to transcription start sites (TSS=position 0; see Extended Experimental procedures). Note that indirect peaks are close to TSS in contrast with the random positioning of reads from control promoters (<10 reads). Only direct peaks harbor the CGATA consensus binding of Beaf32 in contrast to indirect peaks or control sites, as indicated. From these results, we selected peaks based on their intensities and the absence of CGATA motifs within 250 bp windows (see

Extended Experimental Procedures for details). Note that a significant fraction of indirect peaks localize closer (or after/ 3') to TSSs (see text for details).

D. Graph representing the averaged binding profile of Beaf32 in S2 cells as measured by ChIP-Seq (number of reads/ promoter; y-axis) for all promoters harboring no CGATA ('0 CGATA': blue and green curves) or a unique CGATA ('1 CGATA': purple and red curves) and localized close (< 20kbp; green and red curves) or not (> 100kbp; blue and purple curves) to a direct Beaf32 site. Note that this includes all promoters associated with ChIP-indirect peaks or not (see panel C; middle and lower plots) but not direct Beaf32 peaks, resulting in low levels in averaged ChIP-Seq reads per promoter.

Supplementary Figure S2 (related to Figure 2). Hi-C and 3C data over ChIP-indirect peaks

A Graph representing the genome-wide correlation of the binding of Beaf32 in S2 cells as measured by ChIP-Seq (number of reads/ promoter; y-axis) with its binding profile in embryos (Fold enrichment in Log ratio; x-axis).

B. Detection of direct and ChIP-indirect Beaf32 peaks in embryos. The graph shows the relative enrichment (in Log ratio; y-axis) and positioning with respect to TSS (position 0) of control promoter regions (green line), direct (CGATA-containing; red line) or ChIP-indirect (no 'CGATA', blue line) Beaf32 peaks as detected in embryos. The dotted line represent the maximum from each plot obtained.

C. Normalized Hi-C profiles obtained from Hi-C data (Sexton et al., 2012) probing the long-range interactions of distant/direct Beaf32 peaks (> 15 kbp away; orange triangle/green oval) over control- or ChIP indirect peaks (position 0; x-axis), as indicated. Long-range interaction profiles were aggregated over control or ChIP indirect peaks (see legend of Figure 2C; Extended

Experimental procedures). Aggregation was performed for distinct groups of direct Beaf32 binding sites defined as a function of the presence or absence of the overlapping binding sites of the Insulator Binding Proteins (IBPs) dCTCF (yellow and light blue curves, respectively) or GAF (green and blue curves) as compared to all direct Beaf32 peaks (red curve). The relative enrichment in long-range interactions was obtained by normalization over background levels defined by the interactions of the direct peaks with all neighboring bins (enrichment ; y-axis). Note that no significant changes was found for the interactions of direct peaks that harbor no overlapping site for the isoform Beaf32A ('no 32A'; dark blue curve; see text). All data were from embryos (see panel B for detection of ChIP-indirect peaks in embryos).

D. Similar analysis to panel C except that Hi-C profiles were generated for distances > 100 kbp between the direct peaks of Beaf32 and indirect peaks (blue curve) or control neighboring bins (all other curves). See also Figure S7 for Hi-C profiles between sites separated by topological borders.

E. Top: View of the *tsp39D-mio/crc* locus from our Gbrowser showing the ChIP data for Beaf32 (red) and CP190 (green). with a typical direct (> 3 CGATAs) Beaf32 peak in the *tsp39D* promoter and ChIP-indirect peaks in the *mio/crc* promoters. The arrow indicates the orientation of the 3C analysis using two different anchors corresponding to ChIP-indirect peaks (panel F). Bottom: close up view of the 10 kbp region surrounding the 3 typical ChIP-indirect peaks with 0 CGATA (peaks in black or blue) in the *mio/crc* promoters, together with Hind III restriction sites (bottom: vertical bars). Note that one indirect peak harbors a DREF consensus motif ("TATCGATA"; peak in grey) that is otherwise not enriched among ChIP-indirect peaks (< 5 % ; p-value ~1). The horizontal arrows (bottom) show the anchor sites used for the 3C analyses (panel F).

F. Top: Relative frequency chimera products (obtained from 3C data) measured by qPCR analyses in multiple independent 3C assays in Beaf32 (red curve) or CP190 (orange curve) depleted versus control mock depleted (blue curve) cells. The *x-axis* represents the distance from the indicated anchor site (ChIP-indirect peak of Beaf32 ; see blue arrow of panel E). Chimera (ligation product) were measured with a set of primers spanning the whole locus relative to 3 control sites using Taqman-MGB probe (see Extended Experimental procedures). Error bars are standard errors from 3 independent 3C DNA preparation measured in triplicates.

Bottom: Same as top panel except that the relative frequency chimera products were measured with a second anchor corresponding to the left ChIP-indirect peaks ; see black arrow of panel E). Error bars are standard errors from three independent 3C DNA preparations.

Supplementary Figure S3 (related to Figure 3). Synthetic Beaf mutants impair the interaction of Beaf32 with CP190

A. Left panels: Western blotting analysis of nuclear extracts ('NE') using anti-Beaf32 (top) or control anti-histone H3 ('anti-H3'; bottom) antibodies in the presence (+) or absence (-) of the C-terminal peptide competitor also used for affinity chromatography (see Fig. 2), as indicated. Right panels: Western blotting of Nuclear extracts using anti-CP190 antibodies or control IgG.

B. Western blotting following a co-immunoprecipitation experiment with anti-Beaf32 antibodies or control IgG using specific anti- CP190 (top), anti-Beaf32 (middle) or control anti-H3 (bottom) antibodies. Note that both Beaf32 and CP190 are depleted from supernatants ('SUP') and present in eluates ('E') as well as pellets ('P') for precipitations with anti-Beaf32 but not with control IgG.

C. View of the *tsp39D-mio/crc* locus showing the ChIP data for Beaf32 (red) and CP190 (green). The arrow indicates the orientation of the 3C analysis between direct and ChIP-indirect peaks of Beaf32 (bottom panel).

D. Graph showing the binding of WT/mutant -Beaf proteins (as illustrated on the top scheme) measured by qPCR of ChIP samples precipitated with anti-Beaf32 or IgG control antibodies, from RNAi Beaf32 depleted cells stably transfected with WT/mutant -Beaf. The binding levels of WT/mutant Beaf were measured corresponding to the *tsp39D/crc* locus (see arrows).

Supplementary Figure S4 (related to Figure 4). Characterization of synthetic Beaf mutants.

A. Scheme showing the silent mutations introduced into synthetic wild-type and mutated *beaf* (red/green, respectively) by systematic replacement of the third nucleotide ('NNN => NNn') for every available degenerated codon. Note that such silent mutations allowed to express synthetic Beaf from stably transfected cells concomitantly with the RNAi-mediated depletion of endogenous Beaf32 (blue) as shown in panel B and D.

B. Western blotting analysis of extracts prepared from S2 cells transfected with either synthetic wild-type or mutated beaf, before or after RNAi-mediated depletion. The membrane was probed with anti-Beaf32 antibodies recognizing both the endogenous 'Beaf32' as well as the 'wild-type' or 'mutated' synthetic 'beaf' (see upper arrow, 'Beaf-WT/mut') or with anti-histone H3 ('H3') for loading control. Note that in this particular experiment, we used a synthetic *beaf* gene harboring a GFP tag to be able to distinguish on WB WT/mut Beaf proteins from endogenous Beaf32.

C. Wild-type -but not mutant- Beaf specifically interacts with CP190 in vivo. Co-immunoprecipitation experiments with anti-CP190 or IgG control antibodies from extracts

prepared from S2 cells transfected by WT/mutant Beaf32 and depleted of endogenous Beaf32. The membrane was probed with anti-Beaf32 antibodies (exposed ~2 sec.) or with anti-histone H3 ('H3') as loading control (exposed ~ 15 min.). Note that endogenous Beaf32 is not detected under this condition (see panel B).

D. Graph showing the results obtained by RT-qPCR expression analysis of WT or mutated *beaf* in parallel with endogenous *beaf32* as indicated, using cDNAs prepared from stably transfected S2 cells (see Experimental procedures), before ('no RNAi') or after RNAi-mediated depletion of *beaf32* (RNAi).

E. Histogram showing the *in vitro* affinities for specific DNA ('s-DNA'; 3 CGATAs) of mut- compared to WT- Beaf as calculated from fluorescence anisotropy measurements (see panel F; Supplementary Experimental procedures) as previously (Cattoni et al., 2013).

F. Anisotropy fluorescence data comparing binding to specific or non-specific DNA fragments ('S-' or 'NS- 'DNA' ; 3 or 0 CGATAs, respectively) as a function of the concentration of WT/mut Beaf (x-axis; in nM). From such analyses, the calculated Kd (Hill model equation) were = 63 ± 10 nM and =68 ± 5 nM for mut-/WT- Beaf, respectively. Note that Hill equation can not precisely calculate the Kd of non-specific DNA (> 200 nM) in the absence of S-shape curve.

G. RTqPCR analysis in Beaf32 depleted or mock-depleted control cells showing the fold change in expression of genes encoding *Drosophila* IBPs including CP190, dCTCF, GAF, Su(Hw) and Beaf32 (also called '32B'; (Jiang et al, 2009)) or its 32A isoform (Jiang et al., 2009) normalized to controls.

H. Data from RNASeq analysis in stably transfected S2 cells depleted of Beaf32 and expressing the mut- or WT- Beaf. The y-axis represents the fold change in expression of the same genes as in panel G expressed in ratio of RNASeq read counts compared to control (mock-depleted) cells

(y-axis) after normalization to total read counts from three independent RNASeq. Expression of 32A was estimated using reads unique to this isoform (see Extended Experimental procedures).

I. Data from 3 independent RNASeq analyses in stably transfected S2 cells depleted of Beaf32 and expressing the mut- or WT- Beaf. The y-axis represents the % of control or differentially expressed genes depending on the presence or absence in their promoters of a direct, indirect or no peak, as indicated. 'DE': differentially expressed genes ($p < 0.001$); Note that similar results were obtained if pre-filtering among such lists only the genes with a fold change in expression > 1.6x ('DE-2': DE genes with a minimal fold change in expression of 1.6; column 3b; see Extended Experimental procedures).

Supplementary Figure S5 (related to Figure 5). dCTCF depletion impairs the expression of genes associated with dCTCF ChIP-indirect peaks

A. Detection of direct and ChIP-indirect peaks of dCTCF. The graph shows the enrichment (in read counts; y-axis) and positioning with respect to TSS (position 0) of control promoter regions (green line) or direct (dCTCF consensus; red line) or indirect (no dCTCF consensus; blue line) dCTCF peaks as detected in S2 cells.

B. Detection of direct and Indirect peaks of GAF. The graph shows the relative enrichment (in Log ratio; y-axis) and positioning with respect to TSS (position 0) of control promoter regions (green line), direct (red line) or ChIP-indirect (no GAF consensus, blue line) GAF peaks as detected in S2 cells.

C. Data from RNASeq analysis in S2 cells depleted of dCTCF or control cells of the expression of genes encoding *Drosophila* IBPs including CP190, GAF, Su(Hw) or Beaf32 compared to

actin control or to dCTCF in read counts (y-axis) after normalization to the total read counts (see Experimental procedures).

D. Pie charts showing the enrichment of differentially expressed ('DE') genes upon dCTCF-depletion compared to mock-depletion (left) among genes whose promoters harbor a direct ('primary targets') or a ChIP-indirect ('long-range targets) peak of dCTCF or no peak, compared to control genes (right; genes with no changes in expression upon dCTCF-KD). Note that similar enrichment (and %) are found when analyzing up- or down- regulated genes separately.

E. Graph showing the enrichment in misregulated genes in dCTCF-depleted compared to control cells, as a function of their distance to a direct dCTCF binding site (0; x-axis) and the presence or not of an indirect dCTCF peak in their promoters (containing no dCTCF consensus; see Experimental procedure). The y-axis indicate the adjusted p-values (Fisher exact test after Benjamini correction; see Experimental procedures) for the enrichment of the differentially expressed genes as a function of their distance to a dCTCF site, for control (black) or for genes that harbor its indirect peaks (red).

F. Left: Box plot representing genome-wide RNASeq data in cells depleted of dCTCF compared to control cells showing the percentages of differentially regulated genes (%, y-axis) depending on the presence ('ChIP-indirect peaks', red) or not ('control', black) of dCTCF indirect peaks in their promoters. The percentage of long-range targets was calculated among the genes present in 2.5 kbp sliding windows at a 1 kbp up to 40 kbp distance from the closest dCTCF-direct peak (position 0). This allowed to calculate the variations in the percentage of differentially expressed genes among all windows in the interval (see Experimental Procedures). Right: similar analysis for the genome-wide RNASeq data in cells depleted of Beaf32 and expressing either mut- compared to WT -Beaf showing the percentages of differentially regulated genes (%, y-axis)

depending on the presence ('ChIP-indirect peaks', red) or not ('control', black) of Beaf32 ChIP-indirect peaks in their promoters. The percentage of long-range targets was normalized to the total number of genes present in 2.5 kbp sliding windows from 1 kbp up to 40 kbp distances from the closest Beaf32 direct peak (position 0), as done for dCTCF.

G. Heat map showing the enrichment (in Log ratio) of long-range targets depending on the presence in their promoters of a direct Beaf32 or a GAF binding site and/or of a ChIP-indirect dCTCF peak (blue triangles). Also indicated are the enrichment in long-range targets as a function of high or low RNAP II pausing (see Experimental procedures) as previously defined (Gilchrist et al., 2010).

H. Detection of direct and indirect peaks of dCTCF in embryos. The graph shows the enrichment (in read counts; y-axis) and positioning with respect to TSS (position 0) of control promoter regions (green line), direct (red line) or ChIP-indirect (no dCTCF consensus, blue line) dCTCF peaks as detected in embryos.

I. Normalized Hi-C profiles probing the enrichment in long-range interactions between distant ChIP-indirect peaks and direct dCTCF peaks (blue curve). All Hi-C profiles were measured by aggregation plots for distances > 15 kbp between ChIP-indirect and direct peaks (see Experimental procedures). The x-axis shows the Hi-C profiles over +/- 5 kbp surrounding the ChIP-indirect peaks (position 0) or surrounding sites (position 0; all other curves; see Supplementary Experimental procedures). The y-axis shows the enrichment in long-range interactions between the distant direct and indirect dCTCF sites normalized to background (p-value ~ 1e-4). All data were from embryos (see panel H for detection of ChIP-indirect peaks of dCTCF in embryos).

Supplementary Figure S6 (related to Figure 6). Beaf32 influences CP190 binding over promoters of long-range targets .

A. Box plot showing the averaged binding levels of CP190 binding (y-axis) according to the presence or absence of IBPs and/or of direct/indirect peaks. The three left box plots show the CP190 levels according to the presence or absence of a direct or an indirect Beaf32 peak in promoters. The two right box plots show the levels of CP190 binding for genes bound by GAF compared to those that are not (pair-wise Wilcoxon test: p-value 1e-9), as tested in the absence of other IBPs (dCTCF or Beaf32). Note that higher levels of CP190 binding in the presence of GAF is in total agreement with previous results showing that the correlation between CP190 binding and the presence of GAF binding sites (Bartkuhn et al., 2009).

B. Percentages of promoters bound by CP190 corresponding to differentially expressed or control genes that harbor or not of direct- (orange), a ChIP-indirect (blue) peak or no peak (grey).

C. Box plot showing the results from ChIP analysis with anti-CP190 or IgG control antibodies in Beaf32 depleted cells stably transfected with WT/mutant –Beaf (green and orange box plot, respectively). CP190 binding levels were measured by qPCR (see Experimental procedures) corresponding to promoter regions of direct/CGATA-containing Beaf32 peaks (Wilcoxon test: p-value 1e-4). Note that 9/10 of the direct peaks do not overlap with a dCTCF or GAF peak (see text; see Extended Experimental Procedures for a list).

D. Percentages of long-range targets or control genes whose promoters are bound by dCTCF or GAF in the presence or not of a ChIP-indirect Beaf32 peak (blue triangle) peak.

E. Variations in CP190 binding levels in cells expressing mut- compared to WT- Beaf with respect to the distance to a direct Beaf32 site (position 0, x-axis). Variations in binding were measured for each CP190 peak within 2 kbp intervals and normalized to the total number of

CP190 sites within the same window (z-score, see Extended Experimental procedure; y-axis) with a standard deviation in Log p-value from independent replicates (mut- / WT- Beaf).

F. Variations in CP190 binding between cells expressing mut- compared to WT- Beaf (y-axis; in ChIP-Seq reads normalized to IgG control) as a function of high, medium or low RNA Polymerase II pausing. Also indicated are the variations for promoters localized > 50 kbp away ('> 50 kbp') from a direct Beaf binding site.

Supplementary Figure S7 (related to Figure 7). Beaf32 affects the expression of long-range targets as a function of NELF/GAF-mediated RNA Polymerase II pausing

A. Data from RNASeq analysis in S2 cells expressing mutant- Beaf showing the expression of genes encoding NELF-B and -E as expressed in read counts normalized to cells expressing WT-Beaf (y-axis).

B. Box plots representing genome-wide RNASeq data in cells depleted of Beaf32 and expressing either mut- compared to WT -Beaf showing the percentages of differentially regulated genes (%, y-axis) depending on the low or high degree of RNA Polymerase pausing (left panel 'Low'/'High pausing', black and red, respectively; Wilcoxon test: p~1e-21) or on the absence or presence of a GAF binding site (right panel, 'GAF'/'control', black and red, respectively; Wilcoxon test: p~1e-54). The percentage of long-range targets was calculated as previously compared to total genes at the same distances (see Fig S4G).

C. Genome-wide averaged positioning of Beaf32 indirect peaks as a function of high (blue curve; GAF/NELF-bound promoters) or low (red curve) RNA Polymerase II pausing. The positioning of CGATA-containing direct Beaf32 peaks is also indicated (black curve).

D. Normalized Hi-C profiles probing the enrichment in long-range interactions between direct GAF peaks and distant ChIP-indirect peaks of GAF harboring a Beaf32 or a dCTCF site (blue curve; p-value of 1e-3) or no site (grey curves; p-value of 1). All Hi-C profiles were measured by aggregation plots for distances > 15 kbp as previously. The x-axis shows the Hi-C profiles over +/- 5 kbp surrounding the ChIP-indirect peaks (position 0) or surrounding sites. All data were from embryos (see Supplementary Figure S5B for ChIP-indirect peaks of GAF).

E. Top: scheme representing the reporter construct used in which *sop* and *hsp27* promoters are bound by Beaf32 and GAF, respectively. Note that such reporter recapitulates the genomic context in which long-range interactions as detected by Hi-C in the presence of Beaf32/GAF over the ChIP-indirect sites of GAF/Beaf32 (panels D-E), as well as the detection of the ChIP-indirect peaks providing the upstream Beaf32 direct binding site was present (see Figure 4E-F). The arrows represent the location of the primers used to quantify gene expression by RTqPCR (see graph below). Bottom: expression levels of *sop* and *hsp27* as measured by RTqPCR from measures in triplicates normalized to the *neomycin* control in cells stably transfected with mutant or WT -Beaf and/or the presence of GAF in the *hsp27* promoter, as indicated.

F. Normalized Hi-C profiles probing the long-range interactions between direct Beaf32 peaks and distant ChIP-indirect peaks of Beaf32 depending on the presence or not of intervening topological border (Top scheme, +/- green triangle) as previously defined (Sexton et al., 2012). The x-axis shows the enrichment in interactions from Hi-C profiles over +/- 5 kbp around the ChIP-indirect peaks (blue triangle; position 0) in the absence or presence of the intervening TAD border (blue and red curves; p-values ~ 1e-4 and 0.09, respectively) as measured over the same range of distances between direct and indirect peaks (15-60 kbp).

SUPPLEMENTARY TABLES

Table S1 (related to Figure 3): Enrichment in CP190- , GAGA Factor- and dCTCF-binding sites among the 'ChIP-Ind' Beaf32 peaks.

Shown are the percentage, p-values (Fisher exact test) and number (middle) of the Beaf32 indirect peaks (peaks with no CGATA ; see Extended Experimental Procedures for details) that overlap with CP190, GAF, dCTCF, Su(Hw) or Cohesin binding sites. Also shown are the percentages and p-values of these later binding sites that harbor Beaf32 indirect peaks.

ChIP-indirect Beaf32 peaks	CP190	GAF	dCTCF	Su(Hw)	cohesin
% of peaks with a binding site (p-values)	77.4 (1e-300)	20.4 (2e-82)	28.1 (1e-18)	2.1 (0.99)	8.9 (1)
Number of sites	1669	439	407	45	193
% of CP190, GAF, dCTCF, Su(Hw) or cohesin binding sites harboring a Beaf32 ChIP-indirect peak	28.4 (1e-300)	35.3 (2e-82)	24.4 (1e-18)	2.1 (0.99)	10.9 (1)

Table S2 (related to Figure 5): The promoters of up- or down- regulated genes in cell expressing mut- compared to WT- Beaf are highly enriched in CP190 sites.

Shown are the percentage (top), p-values (Fisher exact test; middle) and numbers (bottom) for the promoters of genes that were up/down-regulated in cells expressing mutant-Beaf compared to WT-Beaf and that are bound by CP190- or Beaf32. These data highlight two possible modes of gene regulation by Beaf32/CP190: 1) 'primary targets' defined by the differentially regulated genes that are directly regulated by the binding of Beaf32 to their promoters harboring the *bona fide* CGATAs consensus and a direct Beaf32 peak ~80 reads/promoter on average), in agreement with previous reports (Emberly et al., 2008; Bushey et al., 2009) and 2) 'long-range targets' (see Table S3 for a list) defined by differentially regulated genes distantly localized from such peaks (>1 kbp) and where the binding to promoter regions of Beaf32 is barely detectable as a ChIP-indirect peak (~15 reads/promoter on average) that harbors no CGATA consensus.

	downregulated genes mut- / WT- Beaf	**upregulated genes in mut- / WT- Beaf**
Direct Beaf32 CGATA peaks ('primary targets')	**48.2 (3.1e-51) 543**	**45.9 (6.5e-25) 315**
Beaf32 ChIP-indirect peaks in 'Long-range targets'	**26.8 (7e-29) 302**	**23.6 (1e-100) 162**

Table S3 (related to Figure 5). List of 'long-range targets' with or without ChIP-indirect peaks
The list includes genes that are up- or down- regulated in cells expressing mut- compared to WT-Beaf as identified by RNASeq and that do not harbor a direct CGATA-containing Beaf32 peak in their promoter.

gene	Beaf32 indirect peak	Down-reg	Up-reg	gene	Beaf32 indirect peak	Down-reg	Up-reg	gene	Beaf32 indirect peak	Down-reg	Up-reg
CG11023	0	1	0	vkg	1	1	0	CG31901	0	1	0
l(2)gl	0	0	1	Cg25C	1	1	0	Akap200	1	1	0
CG4822	1	1	0	CG9121	0	0	1	CG31894	0	1	0
CG31975	0	0	1	Msp-300	0	1	0	Sema-1a	0	1	0
CG17075	0	1	0	TotM	0	0	1	Glt	0	1	0
Ipk2	0	0	1	nompC	0	0	1	CG9520	0	1	0
CG5080	0	1	0	CG14005	0	0	1	CG18088	0	1	0
CG14341	1	1	0	Gal	0	1	0	Cks30A	1	1	0
CG14346	0	0	1	CG9098	0	0	1	Aldh	1	1	0
CG17646	1	1	0	CG31642	0	0	1	numb	0	0	1
dpr3	0	1	0	Ddr	0	1	0	CG33298	0	1	0
CG4259	0	1	0	Gef26	0	0	1	CG4389	0	1	0
tho2	0	0	1	CG13982	0	1	0	CG15828	0	1	0
CG7263	0	0	1	Tig	1	1	0	pelo	1	1	0
CG9886	0	0	1	Spn27A	1	1	0	CG31710	1	1	0
dpp	0	1	0	l(2)k09022	1	1	0	Nckx30C	0	1	0
Cyp309a2	0	1	0	xl6	1	1	0	CG13116	0	0	1
CG3246	0	1	0	Pvf2	1	0	1	CG5850	0	1	0
CG17593	0	1	0	CG6630	0	1	0	CG5676	1	1	0
Shaw	0	0	1	TepII	1	1	0	RnrL	1	1	0
bowl	0	0	1	CG13795	0	1	0	cdc2	1	1	0
gene	Beaf32 indirect peak	Down-reg	Up-reg	gene	Beaf32 indirect peak	Down-reg	Up-reg	gene	Beaf32 indirect peak	Down-reg	Up-reg
CG4738	0	1	0	CG16873	0	1	0	CG10343	1	0	1
CG6287	0	1	0	He	0	1	0	CG10623	0	0	1

CG14929	0	1	0	CG8942	0	1	0	CG10431	1	1	0
Pde1c	0	0	1	CG16876	0	1	0	CG10492	1	0	1
CG14933	0	0	1	smi35A	0	1	0	CG10702	0	0	1
CG34163	0	1	0	wb	1	1	0	CG17350	0	1	0
Ced-12	0	1	0	Adh	0	0	1	CG33116	1	1	0
Aats-thr	0	0	1	vig	0	0	1	CG10337	0	1	0
CG6388	1	0	1	TepI	0	1	0	sNPF	0	1	0
CG5792	1	0	1	beat-Ib	0	1	0	fok	1	0	1
MRP	1	1	0	CG4793	0	1	0	CG2493	1	1	0
A16	1	0	1	Ca-alpha1D	0	0	1	CG9338	0	1	0
CG9932	1	1	0	CG5888	1	1	0	Mtp	0	1	0
CG31856	1	0	1	Idgf1	0	1	0	CG9270	0	1	0
CG5142	0	0	1	Idgf2	1	1	0	CG34136	0	0	1
CG16974	1	1	0	Idgf3	1	1	0	CG9248	1	0	1
CG6116	0	0	1	CG13260	0	0	1	Mcm10	0	1	0
CG31729	0	1	0	CG4631	0	1	0	CG8665	0	0	1
Nnp-1	1	1	0	Mhc	0	0	1	CG33813	0	1	0
CG9306	0	1	0	CG13280	0	1	0	CG33902	0	1	0
CG9305	1	0	1	CG4841	0	1	0	CG33817	0	1	0
CG10859	1	1	0	Dif	1	1	0	CG33896	0	1	0
fred	0	1	0	CG7144	1	1	0	CG5188	0	1	0
ed	0	1	0	lectin-28C	0	1	0	CG5322	0	1	0
Sr-CI	0	1	0	CG8486	1	1	0	CG6206	0	1	0
Atet	1	0	1	CG8353	0	1	0	porin	1	1	0
l(2)05714	1	0	1	Btk29A	1	0	1	abo	0	1	0
CG8891	0	1	0	CG7806	1	1	0	RpL9	1	0	1
gene	Beaf32 indirect peak	Down-reg	Up-reg	gene	Beaf32 indirect peak	Down-reg	Up-reg	gene	Beaf32 indirect peak	Down-reg	Up-reg
CG42352	0	0	1	CG1621	0	0	1	Cyp4p1	0	0	1
CG17018	0	0	1	didum	0	0	1	CG13739	0	1	0

CG40040	0	1	0	Corin	0	1	0	prel	1	0	1
Gprk1	1	1	0	scra	1	1	0	CG33757	0	0	1
CG17486	1	1	0	CG30492	1	1	0	CG1888	1	0	1
CG17883	1	1	0	CG30494	0	1	0	cbx	1	1	0
CG17082	1	1	0	CG30493	1	0	1	CG12923	0	1	0
CG12547	1	1	0	CG30496	1	0	1	CG12913	0	1	0
CG14464	1	1	0	Prosalpha6	0	1	0	CPTI	1	1	0
gus	1	1	0	CG11198	0	0	1	CG12325	1	1	0
Nipped-A	1	1	0	Obp44a	0	0	1	CG7686	1	0	1
Ogt	1	1	0	CG8709	0	0	1	luna	0	1	0
CG11665	1	1	0	Cyp6a14	0	0	1	Cyp12d1-p	0	1	0
scarface	1	1	0	CG30357	0	1	0	shn	1	0	1
CG17337	1	1	0	stmA	0	1	0	CG7737	1	0	1
CG11163	1	1	0	CG8740	0	0	1	CG13220	1	1	0
mle	1	1	0	Dmn	0	0	1	CG30022	1	1	0
SCAP	1	0	1	sec31	1	0	1	CG7763	0	1	0
EcR	0	0	1	CG8248	0	0	1	CG7777	0	1	0
Ptr	0	0	1	CG8247	1	1	0	CG9003	1	0	1
l(2)01289	0	0	1	CG8243	1	0	1	Sobp	1	0	1
Spn4	0	1	0	CG8078	0	0	1	CG13185	1	1	0
CG9454	1	1	0	Or-CD2	0	0	1	ERp60	1	1	0
kuz	1	1	0	CG5758	0	0	1	CG33841	0	1	0
CAH1	0	1	0	CG31746	0	1	0	CG33847	0	1	0
CG33307	0	1	0	perd	0	0	1	CG33850	0	1	0
Ance	1	1	0	Cyp310a1	0	1	0	CG33876	0	1	0
CG33115	0	1	0	CG10383	1	0	1	CG15216	0	0	1
gene	Beaf32 indirect peak	Down-reg	Up-reg	gene	Beaf32 indirect peak	Down-reg	Up-reg	gene	Beaf32 indirect peak	Down-reg	Up-reg
CG30046	0	1	0	CG10131	0	1	0	Alk	0	1	0
CG8501	0	1	0	AttB	0	1	0	CG30463	1	1	0

CG8520	0	0	1	CG10249	0	1	0	Chit	1	1	0
CG8525	1	1	0	CG10253	0	0	1	CG5522	1	1	0
CG8545	0	0	1	chn	0	0	1	CG9000	1	1	0
CG8594	1	1	0	igl	0	1	0	fat-spondin	1	1	0
CG8776	0	1	0	CG34123	0	0	1	CG15611	1	1	0
nemy	1	1	0	Mtk	0	1	0	CG11400	0	1	0
Psc	1	0	1	CG34365	0	0	1	CG15917	0	1	0
CG17724	0	0	1	CG8180	1	0	1	CG11395	0	1	0
seq	0	0	1	Flo	1	1	0	snoRNA:U3:54Aa	0	0	1
CG32843	0	0	1	CG8207	0	1	0	snoRNA:U3:54Ab	0	0	1
CG30059	0	1	0	CG30090	0	1	0	mbl	1	1	0
snoRNA:Psi28S-1153	1	0	1	CG33460	0	1	0	Sip1	0	1	0
CG13335	0	1	0	CG30089	0	0	1	CG11419	0	1	0
CG6197	1	1	0	CG30084	1	0	1	CG14483	0	1	0
fas	0	0	1	CG12970	0	0	1	icln	1	1	0
CG30484	1	1	0	CG8314	1	1	0	CG18186	1	1	0
CG16935	1	1	0	SP2353	0	0	1	CG4966	1	0	1
AGO1	0	0	1	Ptp52F	0	1	0	CG6406	0	0	1
CG8067	1	1	0	shark	1	0	1	olf186-M	0	0	1
CG30482	0	1	0	CG15706	0	1	0	CG10915	1	0	1
CG42288	0	1	0	JhI-26	0	0	1	GstE2	1	1	0
O-fut1	0	1	0	CG4398	0	1	0	GstE8	0	0	1
CG8468	1	1	0	CG7848	0	0	1	CG30116	0	0	1
Arc2	1	1	0	Cdk4	0	1	0	CG30115	0	0	1
gene	Beaf32 indirect peak	Down-reg	Up-reg	gene	Beaf32 indirect peak	Down-reg	Up-reg	gene	Beaf32 indirect peak	Down-reg	Up-reg
CG5482	1	1	0	Ugt58Fa	0	1	0	rl	0	1	0
SP2637	1	0	1	CG42260	0	1	0	CG17684	0	1	0
Gint3	0	0	1	nahoda	0	0	1	CG40263	0	1	0

CG15080	0	0	1	CG3520	0	1	0	CG17514	0	1	0
CG15097	0	0	1	CG34372	0	0	1	CG40084	0	1	0
Jheh1	1	1	0	bw	0	1	0	CG41363	0	1	0
Tab2	1	1	0	CG42241	0	1	0	CG7028	0	0	1
CG11055	0	0	1	Psi18S176	1	0	1	CG16971	0	1	0
CG16868	1	0	1	TBPH	1	1	0	CR32477	0	1	0
dpr	1	1	0	Ca-P60A	0	1	0	miple	0	1	0
CG10527	1	1	0	Mlp60A	0	1	0	RhoGEF3	0	1	0
Treh	0	0	1	enok	0	0	1	CG6905	1	0	1
CG9752	1	1	0	Dat	0	1	0	CG3344	0	1	0
Psi18S-1389a	0	0	1	Psi18S-1820	0	0	1	emc	1	1	0
cv-2	0	0	1	CG3376	0	1	0	CG13907	0	1	0
Egfr	0	1	0	nvy	1	1	0	CG34057	0	1	0
CG10433	0	1	0	slik	1	1	0	nerfin-1	0	0	1
CG15678	1	0	1	itp	1	1	0	Glut1	0	1	0
CG9865	0	1	0	CG4806	1	1	0	Rac1	0	1	0
HmgD	0	1	0	CG3880	0	0	1	mtacp1	1	0	1
synj	1	1	0	CG3589	1	0	1	Ptp61F	0	0	1
NtR	1	1	0	ETH	0	1	0	CG9168	0	1	0
CG11275	0	0	1	Ance-5	0	1	0	rho	1	0	1
CG3624	0	0	1	CG30427	0	1	0	drpr	1	1	0
CG6018	0	1	0	CG15861	1	0	1	CG1927	1	0	1
Psi28S-1175a	0	0	1	CG12851	0	0	1	sls	0	0	1
gene	Beaf32 indirect peak	Down-reg	Up-reg	gene	Beaf32 indirect peak	Down-reg	Up-reg	gene	Beaf32 indirect peak	Down-reg	Up-reg
snoRNA:Or-aca1	0	0	1	gsb-n	0	0	1	CG2034	1	0	1
CG32306	0	0	1	RpL18	1	0	1	CG14145	0	1	0
CG16758	0	1	0	CG8607	0	1	0	CG7638	0	1	0
Pxn	0	1	0	CG8602	1	1	0	wls	1	1	0

pgant6	0	1	0	Rac2	1	1	0	GlcAT-P	0	1	0
Shab	0	0	1	CG32373	0	1	0	scyl	0	0	1
snRNA:U5:63BC	1	0	1	CG12262	1	1	0	CG6175	0	0	1
prominin-l	0	1	0	Ect4	0	1	0	chrb	1	0	1
pfk	1	0	1	CG7927	0	0	1	CG18331	0	1	0
CG12034	1	0	1	CG7550	0	0	1	CAH2	0	0	1
CG14961	0	0	1	CG7194	0	0	1	CG32104	0	1	0
CG12014	0	1	0	CG7120	0	1	0	sti	0	1	0
sty	0	1	0	Oseg1	0	1	0	tral	0	0	1
snRNA:U11	0	0	1	Unr	0	0	1	CG10632	1	1	0
CG14985	0	1	0	CG6776	1	1	0	Nplp2	0	1	0
CG1135	0	0	1	rhea	0	0	1	Hml	0	1	0
Chd64	1	1	0	Argk	0	0	1	Rgl	0	1	0
CG1308	0	0	1	Tequila	0	1	0	bbg	0	0	1
dyl	0	1	0	PGRP-LF	1	1	0	CG9238	1	0	1
CG33777	0	0	1	CG3689	1	0	1	CG9425	0	0	1
Myt1	1	1	0	CG32043	1	0	1	mnd	0	0	1
PGRP-LD	1	0	1	CG32036	0	0	1	CG5392	1	1	0
CG32412	1	0	1	LanB2	1	1	0	RhoGAP71E	1	1	0
CG13295	0	0	1	CG18178	1	1	0	comm3	0	0	1
LanA	1	1	0	CG8177	0	1	0	CG6017	1	1	0
CG8398	0	1	0	tna	0	0	1	CG5241	1	0	1
CG9953	1	1	0	CG6321	0	1	0	CG4998	0	0	1
SP1173	1	1	0	CG6404	0	1	0	CG5018	0	1	0
gene	Beaf32 indirect peak	Down-reg	Up-reg	gene	Beaf32 indirect peak	Down-reg	Up-reg	gene	Beaf32 indirect peak	Down-reg	Up-reg
CG4893	0	0	1	CG9619	0	0	1	Ten-m	0	1	0
CG4877	1	0	1	CG9372	0	1	0	Chro	0	1	0
Aats-tyr	0	0	1	Oat	0	1	0	Ssl1	0	0	1
CG32163	1	0	1	Su(Tpl)	1	0	1	CG12768	1	1	0

CG32164	1	0	1	RhoGDI	1	1	0	CkIIalpha	0	1	0
Int6	0	0	1	Ac76E	1	0	1	CG32350	0	1	0
CG9674	0	0	1	CG7668	1	1	0	Snap25	0	0	1
CG32170	0	1	0	fat2	0	0	1	nvd	0	1	0
CG14060	1	1	0	gig	0	1	0	CG40354	0	0	1
CG3764	0	0	1	trbl	1	0	1	CG40178	0	1	0
frc	1	1	0	CG5872	1	1	0	CG40160	0	1	0
CG7603	1	0	1	CG11796	0	1	0	CG40120	0	0	1
Ccn	0	1	0	CG18281	0	1	0	CG41248	0	0	1
Adgf-A	0	1	0	CG4825	1	1	0	CG41262	0	1	0
CG7441	0	1	0	CG4858	1	1	0	CG41274	0	1	0
CG7408	1	1	0	CG13252	1	0	1	Dbp80	0	1	0
Prestin	0	1	0	fng	1	0	1	Gel	1	1	0
W	0	1	0	Ac78C	0	1	0	TwdlG	0	0	1
GNBP2	0	0	1	CG10512	0	1	0	cpx	0	1	0
Cyp12c1	0	1	0	Eip78C	0	1	0	CG11739	1	1	0
CG14073	0	0	1	CG6014	1	0	1	CG32944	0	1	0
l(3)neo26	1	1	0	CG7519	0	0	1	CG34357	0	1	0
CG3961	0	1	0	CG32441	1	1	0	5-HT2	0	0	1
CG11619	0	0	1	CG32446	1	1	0	CG12001	0	1	0
Mkp3	1	1	0	TyrR	0	0	1	CG31531	1	0	1
CG11577	0	1	0	Ddx1	1	1	0	tacc	1	0	1
RpL35A	1	0	1	osk	0	0	1	mbo	0	0	1
CG9629	1	1	0	CG11489	0	0	1	CG11999	0	0	1
gene	Beaf32 indirect peak	Down-reg	Up-reg	gene	Beaf32 indirect peak	Down-reg	Up-reg	gene	Beaf32 indirect peak	Down-reg	Up-reg
Psi28S3091	0	0	1	Prosbeta3	1	1	0	Arp87C	0	1	0
Psi28S3186	0	0	1	pum	1	1	0	Men	0	0	1
CG31548	1	0	1	CG8861	1	1	0	beat-Vb	0	0	1
plx	0	0	1	CG8147	0	1	0	grannysm	0	1	0
CG1239	0	0	1	RnpS1	0	1	0	Ace	0	1	0

CG2091	1	1	0	CG8301	0	0	1	CG9813	0	0	1
CG1218	1	1	0	CG8507	1	1	0	sqd	1	1	0
CG10979	1	0	1	CG3940	0	1	0	CR33328	0	1	0
CG10999	0	0	1	Best1	1	1	0	CG9611	0	0	1
CG9727	0	0	1	Syn	0	1	0	CG3061	1	0	1
dpr11	1	1	0	TfIIFbeta	0	1	0	CG12207	1	0	1
Ama	0	1	0	CG4570	0	0	1	pr-set7	0	0	1
7SLRNA	0	0	1	KP78b	0	0	1	stumps	0	0	1
Scr	0	1	0	CG6744	1	0	1	CG7530	1	0	1
Aly	0	1	0	CG6782	1	1	0	CG7362	0	1	0
CG1939	0	1	0	CG12594	0	1	0	CG14856	0	0	1
gfzf	0	0	1	l(3)neo38	0	0	1	CG3731	1	0	1
CG10050	1	1	0	mthl5	0	0	1	CG31304	0	1	0
snmRNA:331	1	0	1	mfas	1	1	0	CG6912	1	1	0
Sp7	0	1	0	Tk	0	1	0	CG3984	1	1	0
CG2641	1	1	0	svp	0	1	0	CG4334	1	1	0
puc	1	0	1	GstD3	1	1	0	ear	0	0	1
CG18249	0	1	0	GstD7	0	1	0	Aats-ser	0	0	1
CG42286	0	0	1	Cyp9f3Psi	1	1	0	CG5038	0	1	0
CG5399	0	1	0	Cyp12a4	0	1	0	CG5377	0	1	0
CG9797	0	1	0	Aos1	0	1	0	CG18522	0	0	1
pyd	0	1	0	CG6234	0	0	1	CG6045	0	1	0
gene	Beaf32 indirect peak	Down-reg	Up-reg	gene	Beaf32 indirect peak	Down-reg	Up-reg	gene	Beaf32 indirect peak	Down-reg	Up-reg
glob1	0	1	0	CG11779	0	0	1	Dph5	0	0	1
srp	0	1	0	CG7342	0	0	1	CG13843	0	0	1
CG32855	0	0	1	CG6231	0	1	0	Cyp6d4	0	1	0
CG17929	0	1	0	CG4572	1	1	0	CG6985	0	0	1
CG14881	0	0	1	RhoGAP92B	1	1	0	klg	0	1	0

Gyc-89Da	0	0	1	CG12378	0	1	0	CG17119	1	1	0
Gyc-89Db	0	0	1	cic	0	0	1	CG6726	1	1	0
CG18622	1	0	1	CG34118	0	1	0	CG13826	1	1	0
CG8907	0	0	1	Nep4	0	0	1	CG4467	0	1	0
cher	0	1	0	CG17838	0	1	0	DNApol-epsilon	1	1	0
CREG	0	0	1	CG17271	1	1	0	CG10221	0	0	1
Sur-8	0	1	0	CG15695	0	1	0	CG10208	0	0	1
CG5840	1	0	1	Calx	0	1	0	CG6000	1	1	0
AttD	0	0	1	CG10827	0	0	1	sec10	0	1	0
CG7523	1	1	0	Mvl	1	1	0	kal-1	0	0	1
Rim	0	0	1	CG7000	0	0	1	CG6178	1	0	1
cpo	0	1	0	e	0	0	1	CHORD	1	1	0
CG7988	0	1	0	AP-2sigma	1	1	0	cav	0	0	1
cona	0	1	0	CG42335	0	1	0	Acp95EF	0	1	0
fru	0	0	1	CG6637	1	1	0	Orct	1	1	0
fray	1	0	1	CG31431	0	1	0	CG6643	0	1	0
CG14302	0	0	1	CG5740	0	0	1	CG13624	0	0	1
CG7720	0	1	0	PyK	1	1	0	Esp	0	0	1
CG18208	0	1	0	CG13850	1	1	0	bam	1	1	0
CHKov2	0	0	1	CG7950	1	1	0	myoglianin	1	1	0
CG14291	0	0	1	pinta	1	1	0	CG31370	0	1	0
Cyp12a5	0	1	0	Rpn7	1	1	0	CHKov1	1	1	0
gene	Beaf32 indirect peak	Down-reg	Up-reg	gene	Beaf32 indirect peak	Down-reg	Up-reg	gene	Beaf32 indirect peak	Down-reg	Up-reg
CG11902	0	1	0	Tpi	1	0	1	bt	0	1	0
msi	1	1	0	Psi18S-1377b	0	0	1	CaMKII	0	1	0
HLHmbeta	0	1	0	hdc	1	1	0	Rfabg	0	0	1
Lerp	0	0	1	PH4alphaNE2	0	0	1	cals	1	1	0
BM-40-SPARC	0	1	0	tmod	0	1	0	CG11155	0	1	0
sda	1	1	0	CG15547	0	1	0	Kif3C	0	1	0

CG14253	0	1	0	Ptx1	1	1	0	CG33521	0	1	0
NepYr	0	0	1	Sox100B	0	0	1	Dyrk3	1	1	0
Nep5	0	1	0	CG34347	0	0	1	mt:CoI	0	1	0
spz	1	1	0	ttk	1	0	1	mt:CoII	0	1	0
CG14258	0	1	0	CG1896	0	0	1	mt:ATPase6	0	1	0
CG14259	0	1	0	heph	0	1	0	mt:CoIII	0	1	0
eater	0	1	0	CG41329	0	1	0	mt:ND3	0	1	0
CG5646	0	0	1	Gfat1	0	1	0	mt:ND5	0	1	0
CG5611	0	1	0	CR32011	0	1	0	mt:ND4	0	1	0
btz	1	0	1	Ank	1	1	0	mt:Cyt-b	0	1	0
CG4849	0	1	0	CG32000	1	1	0	mt:ND1	0	1	0
Psi28S-3305a	0	0	1	CG2316	1	1	0	mt:lrRNA	0	1	0
Psi28S-3305b	0	0	1	CG31999	0	1	0	mt:srRNA	0	1	0
Psi28S-3405d	0	0	1	yellow-h	0	1	0	cin	0	1	0
inx3	0	0	1	CG1674	0	1	0	CG13377	0	1	0
CG33203	1	1	0	NfI	0	0	1	CG32816	0	1	0
CG14526	1	1	0	zfh2	0	0	1	mod(r)	1	0	1
CG14516	1	1	0	CG1909	0	0	1	CG5254	0	0	1
fig	0	0	1	Eph	1	1	0	Psi18S-531	0	0	1
Cog7	0	0	1	CG1732	0	0	1	CG42248	0	1	0
CG14632	0	0	1	CG2930	1	1	0	CG12065	1	1	0
gene	Beaf32 indirect peak	Down-reg	Up-reg	gene	Beaf32 indirect peak	Down-reg	Up-reg	gene	Beaf32 indirect peak	Down-reg	Up-reg
CG14629	0	1	0	Fas2	0	1	0	AP-1gamma	0	0	1
CG3655	0	1	0	bi	0	1	0	CG12056	0	0	1
CG11418	1	0	1	CG3527	0	1	0	CG12106	0	1	0
CG32813	0	0	1	CG12730	0	1	0	su(r)	0	1	0
CG14786	0	1	0	Tre1	1	1	0	CG34339	0	1	0
pck	1	0	1	CG4096	0	0	1	CG15321	0	0	1

CG14782	1	1	0	mab-21	0	1	0	CG1354	1	0	1
Nmdar2	0	0	1	sqh	1	1	0	CG32698	0	1	0
CG14796	0	0	1	Lag1	0	1	0	l(1)G0230	1	1	0
b6	0	0	1	CG42240	0	0	1	Hk	0	0	1
CG4290	0	0	1	kdn	1	1	0	U3:9B	0	0	1
ph-p	1	0	1	CG3973	0	1	0	spri	1	1	0
CG3835	1	1	0	CG3168	0	0	1	Ork1	0	1	0
Mct1	0	1	0	Psi28S-1060	0	0	1	CG34348	0	1	0
CG2865	0	0	1	CG4586	0	0	1	CG1597	0	1	0
boi	0	0	1	ogre	0	0	1	Spase25	0	1	0
trol	0	1	0	CBP	0	0	1	Kmn1	1	1	0
Syx4	0	0	1	CG32720	0	0	1	e(y)2	0	1	0
CG3603	1	1	0	CG1530	0	1	0	rho-4	0	0	1
rst	0	0	1	CG15332	0	0	1	CG15221	0	1	0
Fcp3C	0	1	0	Tbh	0	1	0	Karl	0	1	0
CG3939	0	1	0	CG18262	0	1	0	CG32663	0	0	1
CG32791	0	0	1	sdt	0	1	0	CG15740	0	1	0
Mnt	0	0	1	Trf2	0	0	1	cac	0	1	0
roX1	0	1	0	CG33223	0	1	0	Tis11	1	1	0
yin	0	1	0	CG12116	0	1	0	CG4330	0	1	0
mew	0	1	0	CG4928	0	1	0	Sep-01	1	1	0
CG1681	1	0	1	wus	1	0	1	CG42343	0	1	0
CG32628	0	0	1	RhoGAP15B	0	1	0	Cyp6t1	0	1	0
gene	Beaf32 indirect peak	Down-reg	Up-reg	gene	Beaf32 indirect peak	Down-reg	Up-reg	gene	Beaf32 indirect peak	Down-reg	Up-reg
ben	0	1	0	CG16700	0	0	1	CG14618	0	1	0
l(1)G0469	0	0	1	snoRNA:Psi28S-2876	0	0	1	l(1)G0196	0	1	0
mRpL38	1	0	1	snoRNA:Psi18S-1854c	0	1	0	CG32499	0	1	0
CG9400	0	1	0	CG42270	0	0	1	CG41476	0	1	0
NetB	0	0	1	CG15816	0	0	1	fog	0	1	0
CG9518	0	1	0	CG12985	0	1	0	CG42346	0	1	0

eag	**0**	**1**	**0**	**CrebB-17A**	**1**	**0**	**1**	**stnB**	**0**	**1**	**0**
rab3-GEF	**0**	**0**	**1**	**CG6361**	**0**	**1**	**0**	**CG9699**	**0**	**0**	**1**
CG15032	**0**	**0**	**1**	**Wnt5**	**0**	**1**	**0**	**CG4768**	**1**	**0**	**1**
CG18210	**0**	**1**	**0**	**CG6891**	**0**	**1**	**0**	**ppk28**	**0**	**0**	**1**
Ac13E	**0**	**1**	**0**	**CG34401**	**0**	**1**	**0**	**CG4829**	**0**	**1**	**0**
CG7860	**0**	**0**	**1**	**Obp18a**	**0**	**0**	**1**	**CG13004**	**1**	**0**	**1**
sd	**0**	**0**	**1**	**CG8034**	**1**	**0**	**1**	**CG4872**	**0**	**0**	**1**
CG32580	**0**	**1**	**0**	**CG8051**	**0**	**1**	**0**	**RhoGAP19D**	**0**	**1**	**0**
CG9911	**0**	**1**	**0**	**gfA**	**0**	**1**	**0**	**CG32521**	**1**	**0**	**1**
sl	**1**	**0**	**1**	**e(y)3**	**1**	**0**	**1**	**CG1494**	**0**	**1**	**0**
Fur2	**1**	**1**	**0**	**CG14225**	**0**	**1**	**0**	**CG1801**	**0**	**1**	**0**
Traf-like	**0**	**1**	**0**	**Bap**	**1**	**1**	**0**	**Rep1**	**0**	**0**	**1**
CG34331	**0**	**1**	**0**	**sol**	**1**	**0**	**1**	**CG8378**	**1**	**0**	**1**
CG9455	**0**	**1**	**0**	**Mys45A**	**1**	**1**	**0**	**CG8878**	**1**	**0**	**1**
CG30157	**1**	**0**	**1**	**CG30344**	**1**	**1**	**0**	**CG8860**	**0**	**1**	**0**
Gadd45	**1**	**1**	**0**	**CG8008**	**0**	**1**	**0**	**CG8306**	**1**	**1**	**0**
cos	**1**	**0**	**1**	**CG8029**	**1**	**1**	**0**				
Cam	**0**	**1**	**0**	**ttv**	**1**	**1**	**0**				

Table S4 (related to Figure 4): The identified dCTCF- / GAF- 'ChIP-indirect' peaks are largely enriched in Beaf32 binding sites. Shown are the percentage (top) p-values (Fisher exact test; middle) and numbers (bottom) for intersections between the direct Beaf32 peaks and the identified ChIP-indirect-peaks of GAF and dCTCF (see also Fig. S5). The right column corresponds to the true ChIP-indirect peaks that harbor no DNA consensus motifs as previously defined for GAF or dCTCF ("no GAF/ dCTCF"; (Negre et al., 2010)).

	Beaf32 direct peaks	**Beaf32 direct peaks (no GAF/no dCTCF)**
GAGA Factor ChIP-indirect peaks	**48.7 (9.3e-31) 318**	**46.1 (2e-32) 301**
dCTCF ChIP-indirect peaks	**40.6 (2.2e-59) 1143**	**38.8 (8.3e-72) 1093**

Table S5 (related to Figure 5): dCTCF-depletion impairs the expression of genes whose promoters harbor dCTCF ChIP-indirect peaks

Shown are the percentage (top), p-values (Fisher exact test; middle) and numbers (bottom) by intersecting data obtained from RNASeq analysis in dCTCF-depleted compared to control cells (see Fig. S5) with promoters harboring ChIP-indirect peaks of Beaf32 (harboring a direct peak of dCTCF) or in the ChIP-indirect peaks of dCTCF (harboring a direct peak of Beaf32 or GAF or both ('all'). No dCTCF control indicates the promoters where no ChIP-indirect peak of dCTCF was detected. Note that genes associated with direct peaks overlapping for two distinct IBPs (e.g. between direct peaks of Beaf32 and dCTCF; Bushey et al., 2009) were excluded to analyze specifically effects through long-range interactions.

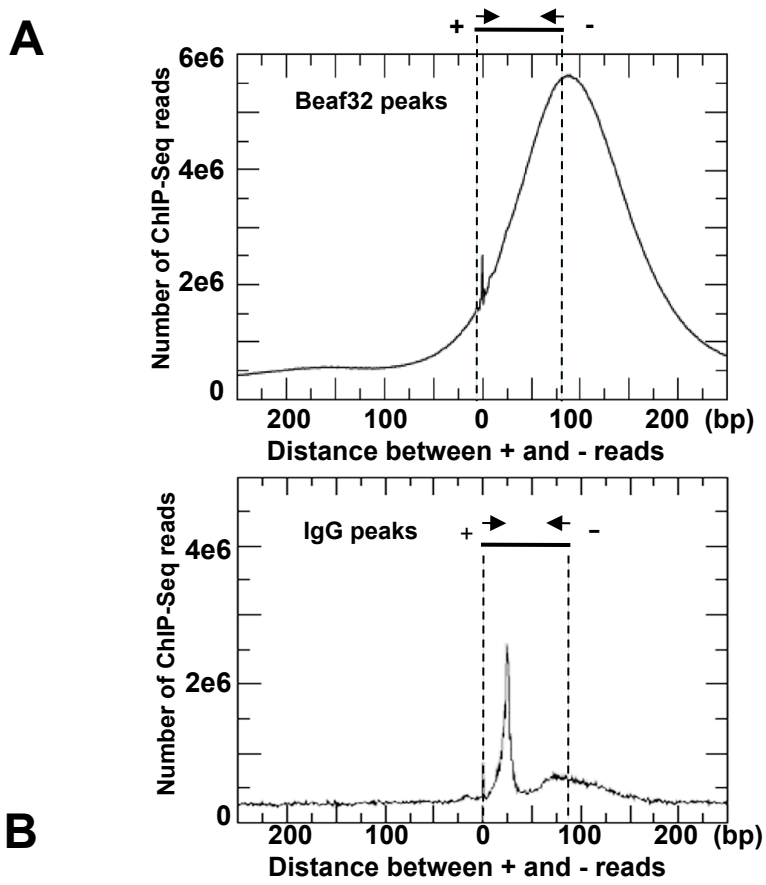
dCTCF depletion ----------	**Beaf32 ChIP-indirect (dCTCF)**	**dCTCF ChIP-In (Beaf32)**	**dCTCF ChIP-In (GAF)**	**dCTCF ChIP-In (all)**	**No dCTCF**
Down-reg	**27.9%** **(1e-31)** **308**	**50.6%** **(3e-59)** **558**	**17.4%** **(5e-17)** **192**	**25.7%** **(1e-4)** **284**	24.2% (1) 267
Up-reg	**22.4%** **(2e-28)** **550**	**44.6%** **(5e-80)** **1096**	**11.8%** **(1e-06)** **290**	**27.1%** **(2e-14)** **665**	24.9% 1 613

Table S6 (related to Figure 7): Enrichment of 'long-range targets' as a function of NELF-mediated RNA Polymerase II pausing

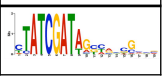
Shown are the percentage (top), p-values (Fisher exact test; middle) and numbers (bottom) for the genome-wide correlation by intersection analyses between 'long-range targets' (see Table S3) harboring Beaf32 ChIP-indirect peaks in their promoters and RNAPII pausing as measured genome-wide by ChIP-Seq (Gilchrist et al., 2010). The number of promoters harboring GAF sites is also indicated.

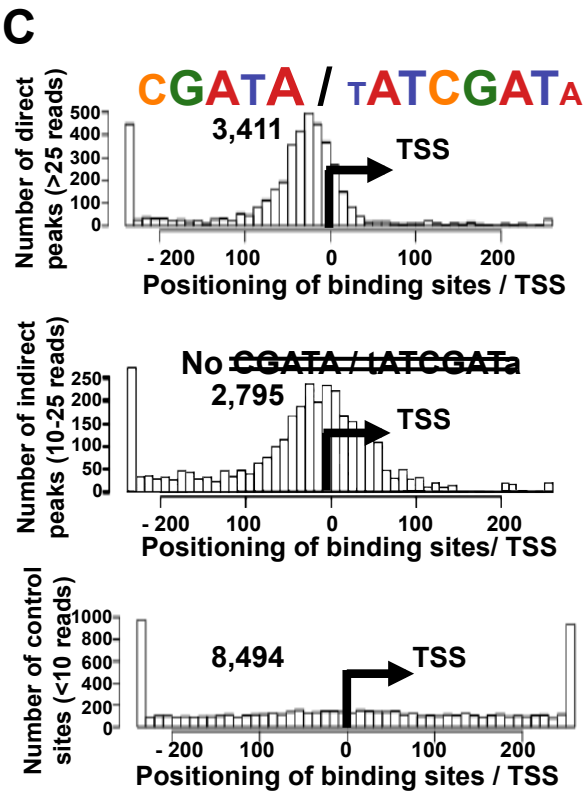
		RNAPII pausing		
		High	**Med**	**Low**
mut/WT -Beaf **Long-range target genes (ChIP-indirect)**	**Down-reg**	**37.2%** **(3e-25)** **77** **(61GAF)**	**19.8%** **(3.e-05)** **41** **(23GAF)**	**14.0%** **(0.05)** **29** **(6GAF)**
	Up-reg	**27.2%** **(2e-13)** **39** **(36GAF)**	**17.3%** **(0.12)** **15** **(9GAF)**	**16.0%** **(0.12)** **15** **(1GAF)**
mut/WT Beaf **control genes (no ChIP-indirect)**	**Down-reg**	**10.7%** **(0.5)** **21**	**8.2%** **(0.9)** **16**	**9.2%** **(0.7)** **18**
	Up-reg	**8.7%** **(0.8)** **15**	**7.6%** **(0.9)** **13**	**12.2%** **(0.2)** **21**
mut/WT Beaf **primary target genes**	**Down-reg**	**26.0%** **(1e-3)** **147**	**24.7%** **(2e-15)** **134**	**22.9%** **(3e-11)** **124**
	Up-reg	**26.9%** **(2e-3)** **85**	**29.5%** **(5e-12)** **93**	**21.2%** **(9e-5)** **67**

Figure S1 Liang et al.



B

Motif	Binding Factor	E-value	Motif
CGATA	Beaf32	1.2e-60	
tATCGATa	DREF	1.2e-59	



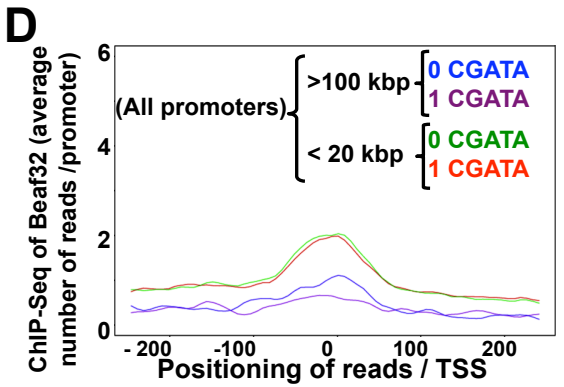


Figure S2 Liang et al.

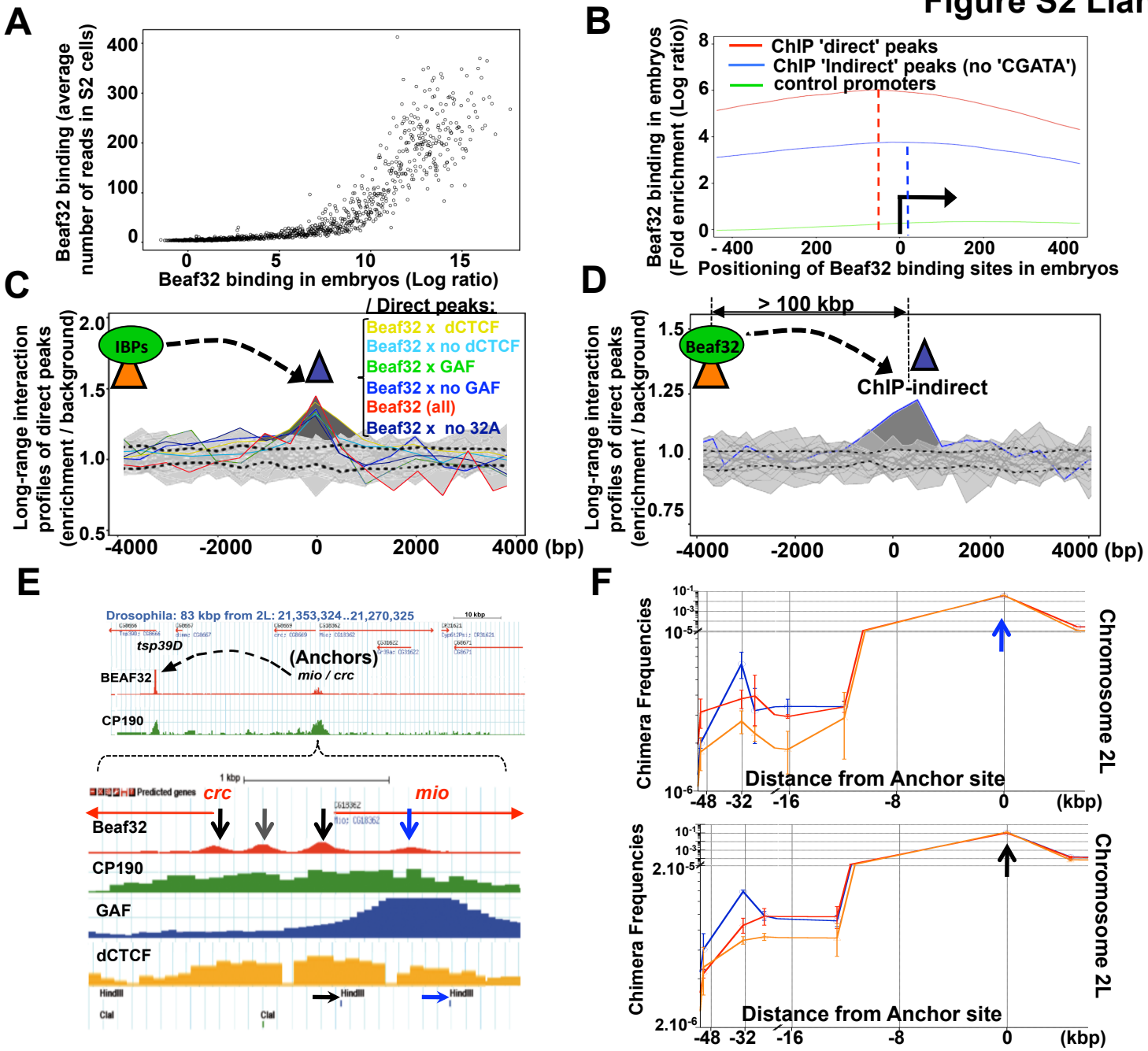


Figure S3 Liang et al.

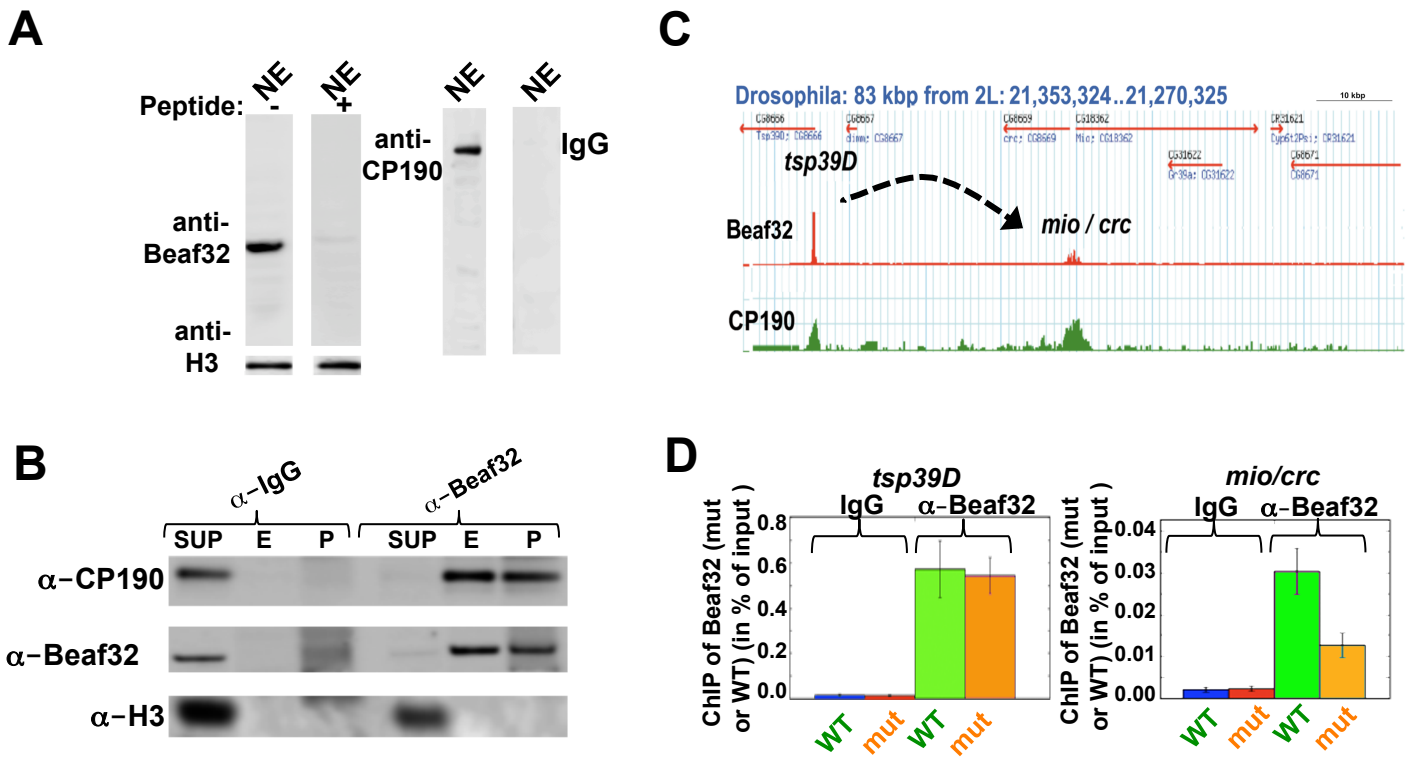


Figure S4 Liang et al.

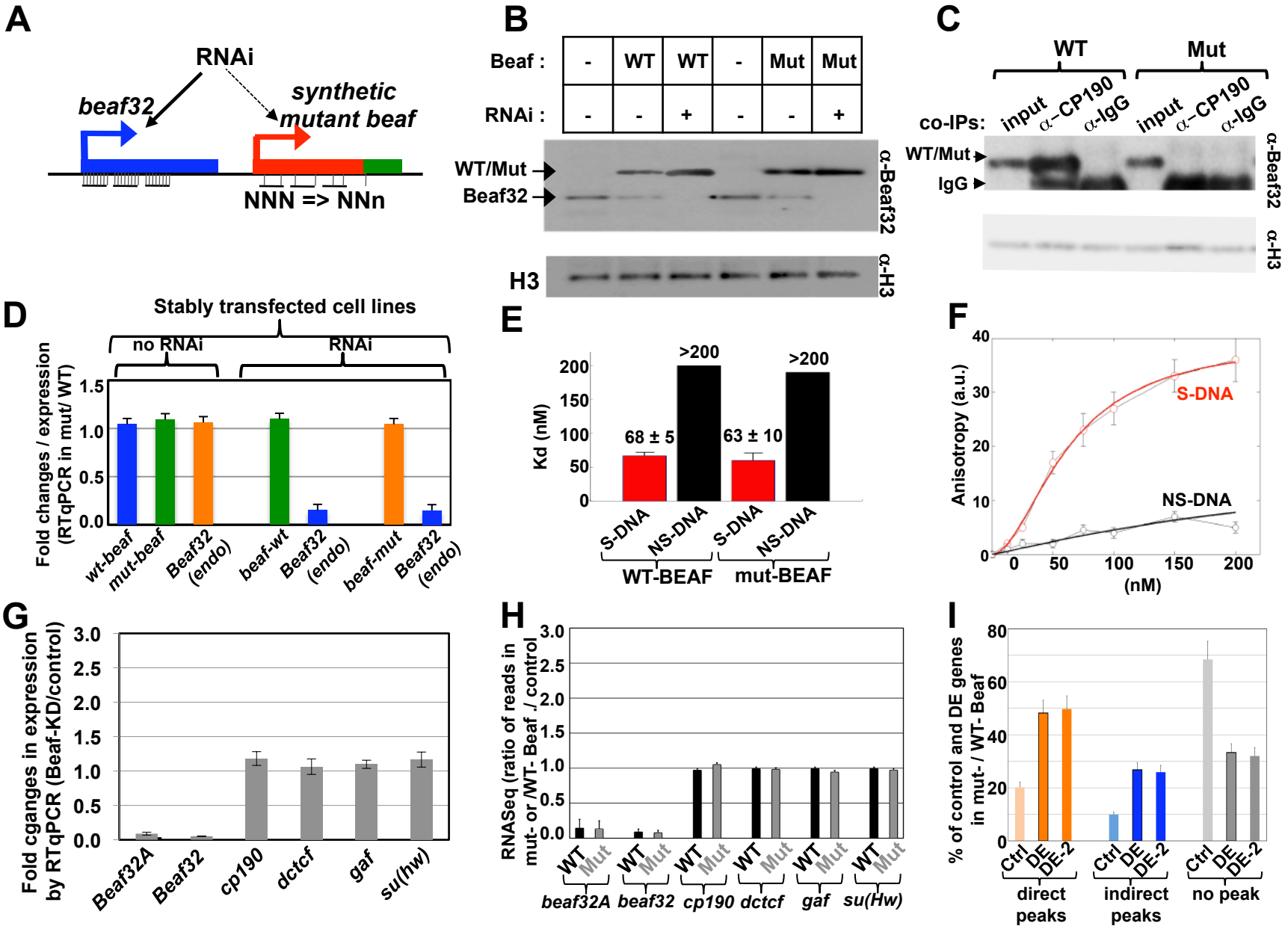


Figure S5 Liang et al.

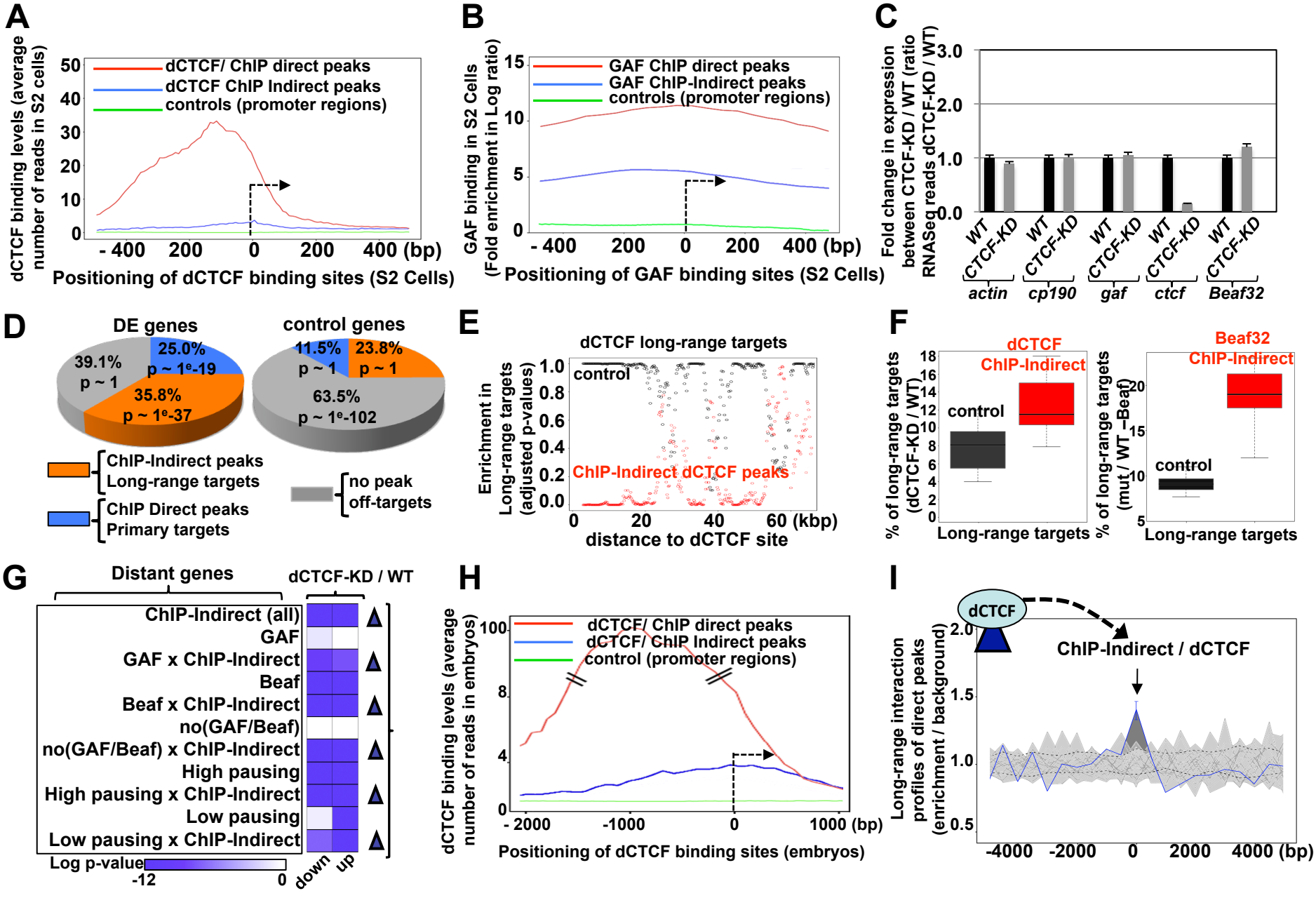


Figure S6 Liang et al.

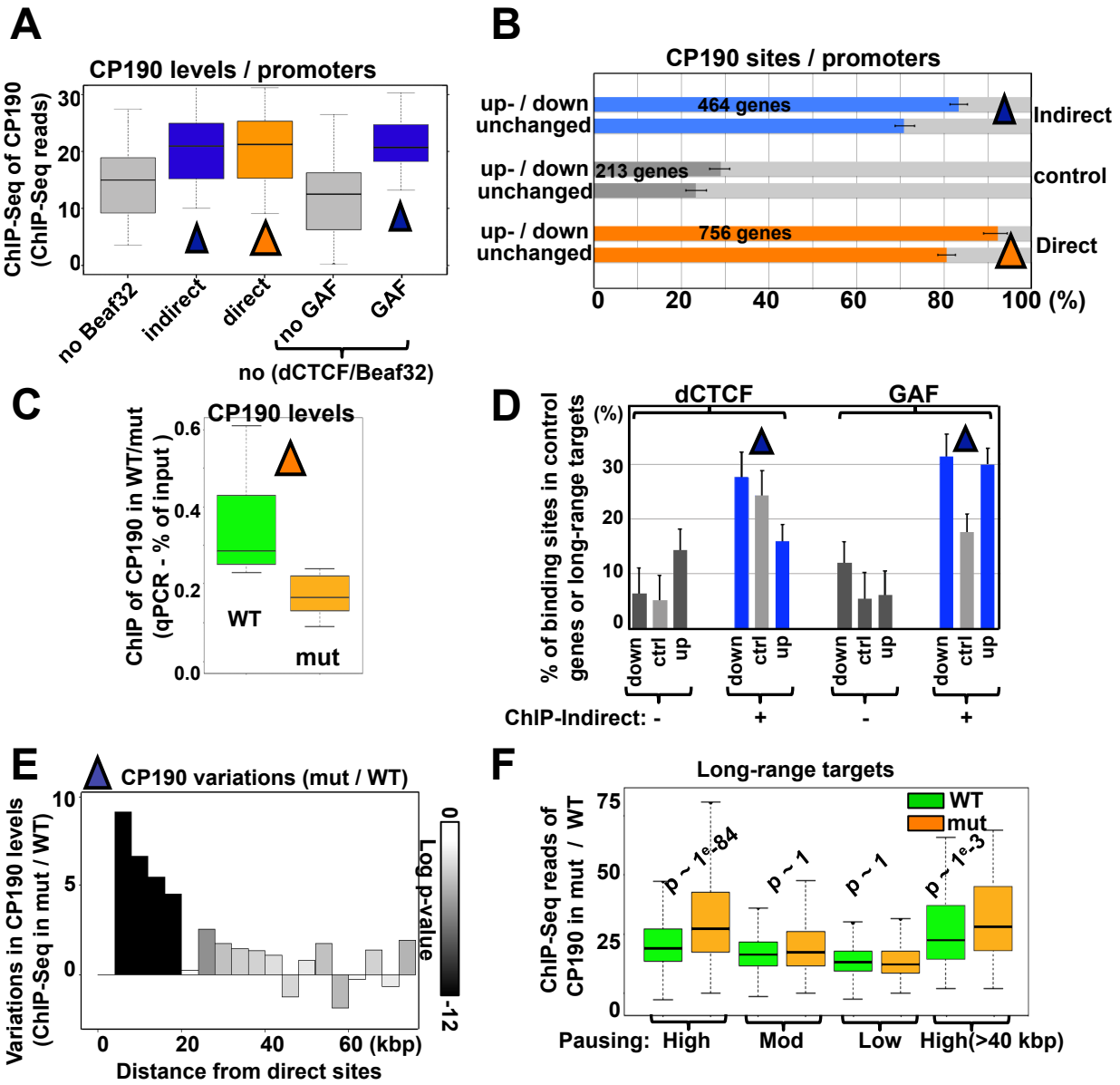


Figure S7 Liang et al.